

12-1990

A Test of a Mastery Learning Approach for Teaching Basic Paragraph Writing Skills to Spanish Language Background Students

Sara Soledad Garcia

Santa Clara University, sgarcia@scu.edu

Follow this and additional works at: <http://scholarcommons.scu.edu/tepas>



Part of the [Education Commons](#)

Recommended Citation

Garcia, Sara Soledad, "A Test of a Mastery Learning Approach for Teaching Basic Paragraph Writing Skills to Spanish Language Background Students" (1990). *Teacher Education, Preparation, and Advanced Studies*. Paper 23.
<http://scholarcommons.scu.edu/tepas/23>

A Dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Education.

This Dissertation is brought to you for free and open access by the School of Education & Counseling Psychology at Scholar Commons. It has been accepted for inclusion in Teacher Education, Preparation, and Advanced Studies by an authorized administrator of Scholar Commons. For more information, please contact rschroggin@scu.edu.

UNIVERSITY OF CALIFORNIA
Santa Barbara

A Test of a Mastery Learning Approach for
Teaching Basic Paragraph Writing Skills to
Spanish Language Background Students

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Education

by

Sara Soledad Garcia

Committee in charge:

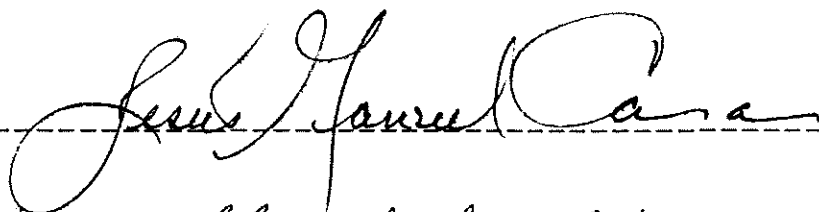
Professor James H. Block, Chairperson

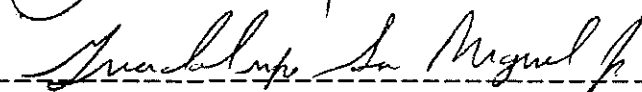
Professor J. Manuel Casas

Professor Guadalupe San Miguel

December 1990

The dissertation of Sara Soledad Garcia
is approved:

A handwritten signature in cursive script, reading "Jesus Manuel Cerna", written over a horizontal dashed line.

A handwritten signature in cursive script, reading "Marcelino La Miguel Jr.", written over a horizontal dashed line.

A handwritten signature in cursive script, reading "James H. Block", written over a horizontal dashed line.

Committee Chairperson

December 1990

ACKNOWLEDGEMENTS

I want to express sincere gratitude to my advisor and dissertation chair Dr. James Block for sponsoring my work, inspiring me to conceptualize my study, believing in my ability to persevere through various stages of the dissertation and providing resources, continual feedback and insight. Thanks also are due to my dissertation committee. My gratitude to Dr. J. Manuel Casas for his guidance, patience, feedback and congenial manner. I acknowledge special gratitude to Dr. Guadalupe San Miguel for the guidance, and the many years of continual intellectual, collegial and moral support.

I wish to express special thanks to Dr. Stanley Nicholson Head Consultant, Instructional Improvement Office, for his consistent advice, guidance, encouragement and friendship throughout the statistical analysis and interpretation of the study results.

I am grateful to the teachers, the principals of the two schools and the students that participated in this study. Acknowledgements are also due to Dr. Delores Austin director of CASE Center for Academic Skills Enrichment and to the Department of Chicano Studies for allocating resources that made it possible to complete this dissertation. Finally, special thanks are due to Dr. Llad Phillips Associate Dean, Academic Affairs, for providing the computer services that were so crucial for the completion of the final document.

This dissertation is dedicated to the
memory of my father,
Agustin M. Garcia,
and to my son's future,
Daniel Agustin Duran

VITA

- March 26, 1948--Born, Villa Lopez, Chihuahua Mexico
- 1973-- B.A. California State University, San Jose
- 1974-- Teaching Credential Cal Lutheran College, Thousand Oaks California
- 1973-1976 Classroom Teacher, Spanish and Bilingual Channel Islands High School
Oxnard Unified School District Oxnard, California
- 1976--1978 Bilingual Program Specialist/Instructor Title VII Resource Training
Center California State University at Fullerton
- 1978--1979 Research Associate, Project BETA Bilingual Evaluation Technical
Assistance, Center for the Study of Evaluation UCLA
- 1979--1981 Field Experience Coordinator, Bilingual Teacher Education, Graduate
School of Education, University of California, Los Angeles
- 1981--M. Ed., University of California at Los Angeles
- 1981--1983 Research Assistant, Reading Comprehension Study, Graduate School of
Education, Educational Psychology, University of California, Santa Barbara
- 1981--1984 Supervisor of Teacher Education, Graduate School of Education, University
of California, Santa Barbara
- 1983--1986 Instructor, Essential Skills, Writing, Reading and ESL Santa Barbara City
College, Santa Barbara, California
- 1985--1987 Writing Skills Counselor, Center For Academic Skills Enrichment (CASE)
University of California , Santa Barbara
- 1986--1988 Teaching Assistant, Department of Chicano Studies University of
California Santa Barbara, California
- 1988--Present Lecturer, Department of Chicano Studies University of California
Santa Barbara, California

PUBLICATIONS

- Glossary of Bilingual Terms* with Robert Cervantes California State Department of
Education, Office of Research and Evaluation, Eric Abstract Series # ED 228845 1982
- "The Dynamic Bilingual Team C Candidate," *Telemetry*, UCLA Teacher Education
Lab. Newsletter, 1981

ABSTRACT

A Test of a Mastery Learning Approach for Teaching Basic Paragraph Writing Skills to Spanish Language Background Students

by

Sara Soledad Garcia

Doctor of Philosophy in Education

University of California, Santa Barbara, 1990

This dissertation tested whether learning could be improved for students of Spanish language background through a Mastery Learning (ML) approach. It was hypothesized that higher levels of achievement and more positive affect could occur in an ML classroom.

Phase I of the study compared two instructional conditions: Mastery Learning and conventional. Because of implementation problems in Phase I, Phase II used a Mastery Learning condition only. In each phase, Pre and Posttests, Formative and Summative Tests, and Affect Questionnaires were gathered from a total of 84 students. The Pre and Posttests were modifications of four Subscales testing General Impressions, General Competency, Paragraph Coherence, Mechanics and two researcher developed Trait Scales, Primary Trait-Main Idea and Secondary Trait-Elaboration. All other tests and questionnaire scales were locally constructed.

Phase I, Pretest findings indicated that ML and conventional students began the experiment roughly equal in achievement. Repeated measures analyses of variance for the Pre and Post Measures indicated

that groups were differentially affected by the treatment across time with conventional group's achievement staying constant on most measures and the ML group's decreasing. The ML group performed lower on the Formative and the Summative achievement tests, too. Affect measures were mixed but favored the ML group.

One major problem with Phase I was its implementation. Under the better implemented ML treatment in Phase II, the gains in student achievement were high from Pre to Posttest. T-test analyses revealed that the ML treatment produced significant gains in achievement especially for Paragraph Coherence, Mechanics, and Primary Trait-Main Idea. The Summative Tests achievement results support the Pre to Posttest gains. Affect measures were less mixed than Phase I and again favored Mastery Learning.

This study has unfortunately not been definitive in testing the viability of using ML to improve learning for Spanish language background students. The study only hints that, when well implemented, ML has promise. More definitive studies must follow. The improvement of learning for Spanish language background students is an issue that simply cannot wait.

TABLE OF CONTENTS

CHAPTER I.....	1
Rationale and Significance of the Study.....	1
Statement of the Problem and Hypotheses.....	7
CHAPTER II.....	13
Research Related to the Study	13
The Classroom Environment and Teacher Attitudes Towards Hispanics	14
Testing, Evaluations and Hispanic Achievement.....	18
Conventional Group Instruction	20
Conventional Classrooms and Intrinsic Motivation.....	27
Summary.....	30
Mastery Learning Structures as an Alternative to Conventional Instruction.....	31
Summary.....	36
CHAPTER III.....	41
Methodology: Original Study Phase I	41
The Research Model	41
Research Hypotheses	43
The Research Design Phase I	45
Developing the Dependent Variable	46
Gauging the First Dimension of the Dependent Variable: Academic Achievement CSE Subscales and Trait Scales.....	47
The Measurement Scales CSE and Primary/Secondary.....	47
Gauging the Second Dimension of the Dependent Variable: Academic Achievement.....	49
Gauging the Third Dimension of the Dependent Variable:.....	50
Motivation	50
Developing the Instrumentation.....	50
Phase I Components.....	50
CSE Scale for Evaluating Writing	50
Primary/Secondary Trait for Evaluating Writing	52
Formative and Summative Measures	53
Formative Test I Part A and B (Phase I).....	54
Formative Test II Parts A, B, C (Phase I).....	55

Summative Test (Phase I).....	55
Questionnaire to Test Motivation.....	56
Developing The Independent Variable	58
The Conventional Instruction.....	58
The Mastery Learning Approach	59
Antecedent Variables.....	60
Prior Achievement	60
Writing Ability	61
Context.....	61
Procedures for Data Collection.....	63
CHAPTER IV.....	69
Methodology: Revised Study Phase II.....	69
Modifications of the Initial Research Design.....	69
The Research Model	70
Research Hypotheses	71
Developing the Dependent Variable	73
Gauging the First Dimension of the Dependent Variable: Academic Achievement CSE Subscales and Trait Scales.....	73
The Measurement Scales CSE and Primary/Secondary.....	73
Gauging the Second Dimension of the Dependent Variable:.....	74
Academic Achievement	74
Gauging the Third Dimension of the Dependent Variable:Motivation.....	74
Developing the Instrumentation.....	75
Phase II Components	75
CSE Scale for Evaluating Writing.....	75
Primary/Secondary Trait for Evaluating Writing	75
Summative Test I Phase II.....	75
Summative Test II Phase II	76
Summative Test III Phase II.....	77
Questionnaire to Test Motivation	78
Developing The Independent Variable	79
The Mastery Learning Approach	79
Antecedent Variables.....	79
Context.....	80
Procedures for Data Collection.....	82
Chapter V.....	86
Data Analysis	86
Scoring, Reliability and Validity Checks.....	87

Pre-Post Test Writing Sample Scoring:	
Procedure.....	87
Establishing Reliability in Scoring	
CSE Subscales and Trait Scales.....	88
Scoring: CSE Subscales	90
Scoring: Primary Trait.....	91
Scoring: Formative and Summative Tests	
Phase I.....	92
Scoring: Formative and Summative Tests	
Phase II.....	93
Scoring: Perception of Achievement,	
Attitude and Interest Questionnaire	94
Coding The Data.....	95
Procedures for Data Coding	95
Coding the CSE and The Trait Scales	96
Coding the Formative and Summative Tests.....	97
Coding The Perception of Achievement ,	
Summary Sample of Variable Codes.....	98
Establishing Reliabilities.....	99
CSE Subscales and Primary/Secondary Trait	
Scales.....	101
Formative Tests Phase I:Formative Tests Item	
Reliabilities.....	101
Summative Test Reliabilities Phase II	104
Reliabilities of the Affect Questionnaire:	
Perception of Achievement, Attitude and	
Interest Questionnaire.....	104
Phase II.....	106
Construct Validity of Measures.....	108
Construct Validity:	
Formative and Summative Tests	115
CHAPTER VI.....	126
Findings.....	126
Phase I Basic Data.....	126
Data Analysis	127
Summary of Repeated Measures	132
ANOVA: CSE Subscales	132
Trait Scales: Primary and Secondary Phase I	134
Summary Repeated Measures ANOVA: Primary	
and Secondary Trait Scales.....	135
Descriptive Results : Perception of Achievement,	
Attitude and Interest Questionnaire Phase I	136

Summary Phase I: Perception of Achievement, Attitude and Interest Questionnaire	137
Formative and Summative Tests Phase I	138
Summary Phase I: Findings.....	139
Phase II Basic Data	140
Data Analysis Phase II	143
Summary Phase II: Perception of Achievement, Attitude and Interest Questionnaire	143
Summative Tests for Phase II	143
Summary Phase II: Findings	144
CHAPTER VII.....	146
Discussion.....	146
Summary of Findings.....	148
Limitations of the Study.....	152
Internal Validity	153
Internal Validity of Phase I.....	153
Internal Validity of Phase II.....	157
External Validity.....	158
External Validity of Phase I (Population).....	159
External Validity of Phase II (Population).....	161
Ecological Validity Phase I.....	162
Ecological Validity Phase II.....	166
Quasi Representativeness Phase I and Phase II	168
The Validity of Mastery Learning Studies.....	170
Conclusions and Interpretations.....	175
Conclusion One.....	177
Conclusion Two.....	179
Conclusion Three	179
Implications for Future Research	181
References.....	186
Appendix A	196
Appendix B.....	224
Appendix C.....	229

List of Figures and Tables

List of Figures:

1. Figure 3.1 Anania' Model	41
2. Figure 3.2 Schema Model of Phase I	43
3. Figure 3.3 Phase I Design	46
4. Figure 3.4 Comparison of Independent Variables	60
5. Figure 3.5 Phase I Procedures Schema	64
6. Figure 4.1 Schema Model of Phase II	70
7. Figure 4.2 Phase II Design	72
8. Figure 4.3 Phase II Procedures	82

List of Tables:

9. Table 5.0 Interrater Reliabilities CSE and Trait Scales.....	90
10. Table 5.1 Samples of Instruments Coding Schemes	991
11. Table 5.2 Internal Reliabilities of CSE and Trait Scales	101
12. Table 5.3 Formative Test 1 Internal Reliabilities	103
13. Table 5.4 Phase I Formative Test 2 Internal Reliabilities	103
14. Table 5.5 Phase I Internal Reliabilities of Summative Test	104
15. Table 5.6 Phase II Internal Reliabilities of Summative Test.....	104
16. Table 5.7 Phase I Internal Reliabilities of Affect I.....	105
17. Table 5.8 Phase I Internal Reliabilities of Affect Scale II	106

18. Table 5.9 Phase I Internal Reliabilities of Affect III	106
19. Table 6.0 Phase II Internal Reliabilities of Affect Scales I, II, and III.....	107
20. Table 6.1 Summary of Instrument Reliability.....	107
21. Table 6.2 Intercorrelations of CSE Subscales	112
22. Table 6.3 Intercorrelations Primary and Secondary Trait Scales.....	114
23. Table 6.4 Intercorrelations of CSE Subscales with Primary and Secondary Trait Scales.....	115
24. Table 6.5 Phase I Principal Components for Formative and Summative Tests Control Group.....	1161
25. Table 6.6 Phase I Principal Components for Formative and Summative Tests Treatment Group.....	120
26. Table 6.7 Phase II Principal Components for Summative Tests	1221
27. Table 6.8 Summary of Instrument Validity.....	1251
28. Table 6.9 Phase I Overall Means and Standard Deviations for CSE and Trait Scales	1261
29. Table 7.0 Repeated Measures ANOVA CSE Subscale 1.....	1281
30. Table 7.1 Phase I Overall Means Dependent Variable CSE Subscale I	128
31. Table 7.2 Repeated Measures ANOVA CSE Subscale 2.....	129
32. Table 7.3 Overall Means Dependent Variable CSE Subscale 2.....	129
33. Table 7.4 Repeated Measures ANOVA CSE Subscale4	130

34. Table 7.5 Overall Means Dependent Variable CSE Subscale 4.....	130
35. Table 7.6 Repeated Measures ANOVA CSE Subscale 6.....	131
36. Table 7.7 Overall Means Dependent Variable CSE Subscale 6.....	131
37. Table 7.8 Repeated Measures ANOVA Primary Trait.....	133
38. Table 7.9 Overall Means Dependent Variable Primary Trait	133
39. Table 8.0 Repeated Measures ANOVA Secondary Trait.....	134
40. Table 8.1 Overall Means Dependent Variable Secondary Trait	134
41. Table 8.2 Phase I Means Repeated Measures Affect Questionnaire	136
42. Table 8.3 Means and Standard Deviations Summary of Phase I Formative and Summative Tests	138
43. Table 8.4 Phase II Means and Standard Deviations of CSE and Trait Scales.....	140
44. Table 8.5 Phase II Matched Case t-Tests.....	140
45. Table 8.6 Phase II Means of Repeated Measures of Affect Questionnaire	142
46. Table 8.7 Phase II Summary of Summative Tests Results.....	144
47. Table 8.8 Summary of Observed Empirical Relationships.....	151
48. Table 8.9 Summary of Study's Limitations	170
49. Table 9.0 Summary of Kulik and Kulik, Banger-Downs (1990).....	173

December 1, 1990

Copyright by
Sara Soledad Garcia
1990

CHAPTER I

Rationale and Significance of the Study

Recently there has been an increasing concern among educators about the growing rate of functionally illiterate young American adults. As defined by Johnson (1975), functional illiterates are people who are unable to perform necessary daily tasks (such as reading a menu or bus schedule, balancing a checkbook or filling out a job application) because of their inability to read and write. Functionally illiterate adults cannot contribute as active citizens to a highly technological society where anything above the most menial of tasks requires expression in writing. Their employment opportunities are usually limited, they are unable participate in the democratic process as informed voters, and they cannot help their children escape a comparable fate (Kozol, 1985). These individuals are bound to be economically, politically, and socially exploited.

Of special concern to educators has been the fact that many functional illiterates come from minority groups. Over a decade ago Johnson (1975), for example, found that more than 40 % of the blacks and Hispanics¹ he surveyed were functionally illiterate. Also of

¹ The term "Hispanic" used in this document appears interchangeable with the terms "Mexican American" and "Chicano". Typically the term Hispanic includes all Spanish surname individuals living in the United States (i.e., Puerto Ricans, Cubans etc.). This study focuses only on the Mexican American population of California and thus refers to this group as Hispanics when data appear labelled as such in the literature reviewed, and in other sections of the document as "Mexican" or "Chicano". These terms as in Laosa's article (1982) refer to persons who trace their lineage to Hispanic or Indo-Hispanic forebears who resided within Spanish or Mexican territory that is now part of the southwestern United States. As Laosa states it also refers to persons who were born in Mexico and

concern to educators has been that more minorities tend to drop out of school than young white people (Camp, 1980). Though Hispanic young people represent only 21 % of the youth population, they accounted for 23 percent of the out-of-school youth (Rumberger, 1982, 1983). Recently Ernest Boyer, President of the Carneige Foundation for the Advancement of Teaching, stated during a live television presentation, that half of the Hispanics in this nation drop-out of school (Public Broadcasting Station, 9-4-86).

There appears to be a strong correlation nationally between the number of minority illiterates and the severe drop-out rate of minorities. Indeed, this correlation between illiteracy and dropping out of school has long been documented. For example, Dentler and Warshauer (1968, 1965) found that illiteracy correlated high with dropping out of school ($R=.84$ for white dropouts and $R=.91$ for nonwhite dropouts). Similarly, Galloway (1985) found in a sample of children selected for a study on absenteeism that 50% were at least two years behind in reading skills. Schreiber (1964) contends that such reading retardation frequently leads to grade retardation which in turn results in dropping out of school completely.

The correlation between minority illiteracy and dropping out may be even stronger in states with a large and growing minority population. In a recent report prepared by the State of California Assembly Office of Research (1985), for example, the California State Library estimates that 2 to 4 million adults are functionally illiterate and that many of these illiterates are Hispanics who dropped out of high school. Two major urban California school districts have published detailed reports which show that in Los Angeles the Hispanic dropout rate was 70 % higher than that of white students, in San Diego it was almost twice that of white students (Camp, 1980).

now hold permanent residency in the United States or whose parents or ancestors immigrated to the United States from Mexico.

During the last decade the causes for dropping out of school have been extensively researched. Yet controversies regarding the educational experience center on questions about the consequences of schooling and about the desire to find the causes, correlates, or motives underlying the actions of dropouts (Casas and Furlong, 1986; Laosa, 1984; Wehlage and Rutter, 1986). Few research studies on dropouts focus specifically on within group factors showing why Mexican Americans do poorly in school or take into consideration the diversity of the subgroups which comprise the Mexican American population in the United States (Casas and Furlong, 1986). Most studies simply underscore the fact that the dropout rate of Hispanics (all Spanish speaking groups) in comparison to whites is dramatic.

Casas and Furlong's (1986) review of the research is the rare exception. To better understand the dropout problem and how it specifically relates to Mexican American students, Casas and Furlong reviewed individual factors, home family factors, and school community factors that result in dropping out of school. They contend that "these three areas directly and indirectly, collectively and interactively contribute to the prevailing dropout problem found in the Mexican American community." Of particular relevance to the present study are studies reviewed by Casas and Furlong related to individual and school factors.

Many researchers consider the possibility that certain student characteristics in combination with certain school conditions are responsible for students' decisions to leave school early. Steinberg, Blinde, and Chan (1984), for example, have reviewed studies explaining the dropout rate of language minority youth--youth identified by their national origin and their degree of non-English language acquisition. These researchers propose that even when family factors such as number of parents, number of siblings, and amount of reading material in the home are controlled, Hispanic youngsters still drop out at a rate far in excess of either white or black

students. Aptitude variables such as reading and math aptitude also contribute to leaving school above and beyond the contribution of family and social class as do self-esteem variables. It seems that potential dropouts score lower on measures of self esteem than those who eventually graduate (Blinde et al., 1984).

Since so much of the literature tends to ignore these students' characteristics by school process interaction and to focus on the student characteristics alone, Steinberg, et al. (1984) have called for much more extensive research on the characteristics of schools. They state that many writers have suggested a discrepancy between the demands and expectations of schools, which tend to reflect the dominant middle class ideology, and the socialization and preparation of minority and lower class youngsters. According to these researchers, it is well documented that minority and lower class youngsters perform more poorly in school than do white and middle class youngsters. But it is not well documented whether this performance differential is due to a lack of fit between youngsters abilities and needs and their schools' demands and resources. For example, it is known that teachers and school personnel are more likely to interact negatively with lower class, minority, and non-English speaking youngsters than with children of their own racial background (as documented by Laosa, 1977). But it is not known whether these negative interactions precede, catalyze, or follow students' poor academic performance.

Some negative interactions seem to precede poor academic performance. For example many studies conducted in the seventies and eighties (Alexander, Fennessey, McDill, and D'Amico, 1979; Oaks, 1985; Rist, 1970; Rosenbaum 1976, 1978; Treadway, 1985) indicate that tracking mechanisms serve to restrict the opportunity of those students who are placed in the lower tracks. Steinberg et al. (1986) state that the educational choices of students in the lower tracks are often based on minimal, erroneous, or inadequate

information. Furthermore, they contend that it can be assumed that choices concerning leaving school for the students in the lower tracks are similarly misguided.

Other negative interactions seem to catalyze poor academic performance. For example, the ambiance of the school classroom has been seen as catalyzing negative interactions for academic failure or success of students. Steinberg, et al. (1986) state that it is widely believed that the nature of the school experience of non-English speaking students is significantly different from that of their peers. Consider the issue of instructional responsiveness. Hernandez (1973) has concluded that one of the main factors affecting Mexican American achievement is the lack of instructional responsiveness on the part of the educational system to the needs of Mexican Americans and other minority groups. She states that little individualization of instruction is reported to take place during the schooling of minority youngsters and suggests that many schools have inflexible curricula that do not account for different levels of student readiness, little variation in approaches to teaching English, and personnel who view cultural or languages differences as deficiencies.

These researchers surmise that language minority youngsters, especially Hispanics drop out of school at a higher rate than their English-speaking peers primarily because language minority youngsters are more likely to encounter academic difficulties early in their scholastic careers. Early Wehlage and Rutter's (1986) research may be extrapolated to suggest that such difficulties pose a serious problem with the holding power of school for these youth. Dropouts, they contend, do not expect to get as much schooling as their peers. They do not perform as well as their peers on school tests, their grades are lower than those of their peers, they are more often truant both in and out of school, and generally get into more disciplinary trouble than other students. Furthermore,

their data suggests that school factors related to discipline are significant in developing a tendency to drop out. Finally, Wehlage and Rutter contend that the process of becoming a dropout is complex because the act of rejecting an institution as fundamental to society as school must also be accompanied by the belief that the institution has rejected the person. The process is probably cumulative for most youth. It begins with negative messages from the school concerning academic and discipline problems. These researchers feel that while most of the literature on dropouts is directed only at deficiencies found in the marginal student, those same characteristics are a reflection on the institution.

As a final point, Wehlage and Rutter state that some negative interactions seem to follow from academic failure. Although language minority Hispanic youngsters drop out at a higher rate than do language minority youngsters from non-Hispanic backgrounds, some studies indicate that non-English-speaking youngsters whose primary language is not Spanish suffer little educational disadvantage.

In summary, it appears that the problems associated with illiteracy and dropping out of school are very complex. For the Mexican American students in the public schools, particularly in states such as California, these problems have reached unparalleled proportions and this has broad implications for the social and economical mobility of future generations. Although recently (May, 1990) the California State Department of Education released a report that the state dropout rate had declined by 18% during a three year period, the battle to reduce attrition is by no means won. A printout of districts' dropout rates was circulated throughout the state. These figures were derived by comparing the accumulated loss of students (10th through 12th grade) from the same graduating class, from 1986 to 1989. The dropout data were based on information submitted by the California Basic Educational Data System each

October. However, according to San Miguel (1986) state reports issued by various private and government groups have publicized the lack of educational opportunities for disadvantaged children and youth and conclude that the problem of dropouts and underachievement continues to be a serious one, especially among language minority groups. Given this fact, a more specific examination of the problems and additional empirical data are needed in order to find effective interventions that will promote excellence in learning for all groups, but especially for the diverse groups which comprise Mexican Americans.

Statement of the Problem and Hypotheses

While some researchers present a dismal picture of the personal characteristics of students that dropout (Beck and Muia, 1980), Wehlage and Rutter (1986) state that it is not clear if measured characteristics such as low educational/occupational aspirations, weak sociability, negative school attitudes, low self esteem and external sense of locus of control are brought to the school or produced by school experiences. They contend that the research on dropouts continues to focus on the relatively fixed attributes of students, the effect of such research may well be to give schools an excuse for their lack of success with the dropout. Furthermore, they state that since traditional research has tended to identify characteristics least amenable to change, the focus of new research might better be directed toward understanding the institutional character of schools and how this affects the potential dropout. These researchers feel that by searching for school factors that contribute to marginal students' decisions to drop out, such research can provide grounds for school-based reform. Although schools can do

nothing about students SES or innate ability, important contributing factors to the dropout rate that are under the control of the school may be modified to change the school conditions of students at risk. These factors include, but are not limited to, teaching practices, low teacher expectations, irrelevant and rigid instruction, tracking, and the use of culturally biased tests, toward Hispanics which engender a negative school environment.

The interrelated issues concerning leaving school early are strongly polarized between the student or potential dropout and the system of schooling. The schooling system, in terms of propagating factors which will promote negative interactions, is on one side and the characteristics of the individual student on the other. The dynamics of the interrelatedness of these factors make it difficult to conduct analysis since not enough research has been conducted on specific dimensions of the dialectics between the individual and the various aspects of the process of schooling.

Let us consider, for example, the impact of teaching and other educational practices. More specifically Erikson, (1987) states that cultural difference can, for a variety of reasons, be an initial source of trouble between teacher and students. Moreover, he states that "what may have begun as simple misinterpretation of intent and literal meaning can develop across time into entrenched, emotionally intense conflict between teacher and student" (pp. 348). The cycle has been observed to be repeated from year to year during elementary school (McDermott, Gospodinoff, 1979).

Erickson concludes that teachers and students in such regressive relationships do not bond with each other. Mutual trust is sacrificed and over time the students become alienated from school. It is no longer just a matter of difference between teacher and student that derives from intergenerationally transmitted communicative traditions, it is also a matter of cultural invention as a medium of resistance in a situation of political conflict. Furthermore, he

contends that as students grow older and experience repeated failure and repeated negative encounters with teachers they develop oppositional cultural patterns as a symbol of their disaffiliation with what they experience (not necessarily within full-reflective awareness) as an illegitimate and oppressive system.

What appears to happen to Mexican American minority students in environments in which repeated academic failures occur is that they develop a feeling of futility and incapability. This process stifles learning and their academic performance begins to weaken. They do not perform well on academic tasks or on standardized tests, since these tests are culturally biased and thus not relevant to the students life experience or knowledge. For the teacher this creates a situation where their expectations of these students diminish and what eventually follows is the application of tracking mechanisms that place low-achieving students into segregated ability groups or classrooms.

The student's perceptions of their academic self esteem are affected by all educational practices that occur in a school setting. If students experience continual failure and teachers lower their academic expectations of these minority students the sense of trust between student and teacher is weakened. Students start feeling a sense of helplessness since they feel disempowered. Motivation for learning decreases and what may occur instead is that these students, in order to protect the sense of self worth, begin using self-serving excuses or manifest other types of behavior in order to survive in these advers classroom environments (see Bricklen and Bricklen, 1967; Covington, 1986; Covington and Omelich, 1979). Carter and Segura (1979), contend that "if the environment regularly and continually presents tasks beyond the person's ability to deal with them, a negative view of self develops. Negative self concept of ability discourages achievement" (p. 236). These doubts reinforce their negative attitude toward school.

Suarez-Orozco (1984) contends that because of the inequitable atmosphere that minority students come to experience from formal schooling, they begin to see education not only as irrelevant, but worse--the traditional educational system, run by mainstream culture, becomes psychologically "a threat" to one's sense of ethnic belonging. When schools become a stage for enacting the inequality and depreciation in the encompassing social structure, success in school may induce what DeVos (1978) has termed a state of "affective dissonance." Essentially these students "learn not to learn."

Considering that the structure of the conventional classroom environment has traditionally failed a great number of students and promotes lower academic expectations on the part of the teacher toward these students, a closer examination of how this environment affects the academic success of minority students is warranted. Teachers cues and positive reinforcement in a conventional environment tend to be directed toward students who are most able to benefit from the instruction--only the high achieving students. Brophy and Good (1976), concluded that observed teacher behavior was consistent with the hypothesis that differential teacher expectations function as self-fulfilling prophecies.

Conventional classroom environments also tend to be competitive in nature. In such environments achievement criteria are primarily normative, (i.e., students are judged relative to their position within the group) the student must compete with others to determine his relative group standing. While competition may spur some students, much of learning and development may be destroyed if competition is the primary basis for motivation.

In addition to competition under a conventional instruction mode, tracking mechanisms play a prominent role. What may be just as important as tracking's impact on student participation is its impact on students feelings of self worth. Students channelled into low tracks start developing doubts about their own ability. The

negative effects of tracking have been documented by many researchers (Carter and Segura, 1979; Loasa, 1981; Orfield, 1980). Students who have been tracked into lower academic tracks are aware that they are differentially treated. Student's perceive it as being even more futile to put forth any effort in attempting to cope with the school environment. The feeling of inadequacy in coping with school demanded tasks limits ability to perform them in the future.

The use of standardized tests within conventional education environments is viewed by many not as a legitimate tool for evaluating learning patterns but as a means of keeping students at lower levels. It has long been documented that minority students perform lower on standardized tests than mainstream students (Duran, 1986; Carter and Segura, 1979; Sanchez, 1932, 1934). Reasons associated with lower performance on these tests are that these tests are culturally and linguistically biased toward minority students. Most important is the fact that these tests do not reflect the classroom instructional objectives and are irrelevant to what is being learned in school.

The conventional schooling practices enumerated act in combination and result in non-equitable opportunities for most students. These continual practices have greatly depressed the achievement of elementary and secondary Hispanics and help generate a negative school experience (Carter and Segura 1979; Erickson 1987; Steingberg et al., 1984). In turn, they have no doubt contributed to the declining enrollment of these students at the university or college level (Duran, 1983; Ponterotto and Casas, 1986). Thus, the cycle continues: underrepresentation in the state university systems means fewer Hispanics with college degrees, and therefore, more Hispanic representation in the labor force and in the State's social service lines. This denies the future generation of intergenerational mobility so crucial to social and economic access.

"Higher educational attainment for Hispanics increases their chances of higher income levels; higher income levels for Hispanics increase the chances of a better economic situation for all citizens" (Rodrigues, 1985, p. 3).

This study tested one educational method of breaking the vicious economical, social, and political cycle into which Hispanics have fallen. The intent of this study was to examine the degree to which a "Mastery Learning" approach, as an equity structure, is a better approach for teaching and learning than a conventional approach for students of Hispanic background. The focus of a Mastery Learning program is on the importance of sufficient rewards for all students, irrespective of ability level. Mastery Learning studies propose that interest and attitude toward learning can be altered positively for all students regardless of placement or prior achievement. If Hispanics are taught by mastery methods they may acquire those interests and attitudes instrumental for staying in school. In this study, therefore, the cognitive and affective learning of Hispanic students will be investigated under two learning conditions: conventional group instruction and Mastery Learning.

CHAPTER II

The educational record for students of Mexican descent in the Southwest is and has been dismal. Carter and Segura (1979) surmise that the compensatory education programs designed to alleviate the conditions of academic failure for Chicanos have not worked. Most schools provide few intrinsic rewards for Chicano students, and the improvement of the school social climate through elimination of negative conditions enumerated above is rarely, if ever, seriously attempted (Carter and Segura, 1979). Teacher behaviors, classroom environment, and institutional procedures alluded to in the previous chapter have further contributed to a negative educational experience for Hispanics. According to some researchers the competitive atmosphere of the classroom and ability grouping seem to be major contributors to the dismal educational record of Hispanics (Carter and Segura, 1979; Duran, 1983; Ogbu, 1978). Tracking and ability grouping, negative and low teacher expectations, negative attitudes and biased testing and evaluation procedures also alluded to in the previous chapter have also contributed to this dismal record.

Research Related to the Study

Recognizing the varied contributing factors that have negatively impacted the educational record of Mexican American children, the theoretical literature reviewed in this section spans several areas. The first section treats the literature on teacher attitudes and expectations as they relate to Hispanic achievement and motivation that is relevant to the educational experience. Then a discussion of testing and evaluation follows. Finally, special

emphasis is given to how conventional classroom environments, which stress competition, diminish motivation for learning. The notions of motivation considered in this review include students' conception of effort and ability, attributions for success or failure, self-worth theories as they relate to teachers expectations, and evaluations, testing, and conventional competitive environments. In the final section studies on Mastery Learning are examined and the applicability of this learning approach as an equity structure for development of higher achievement and better perceptions of achievement for Hispanic students is considered.

The Classroom Environment and Teacher Attitudes Towards Hispanics

It has been continually stated by researchers and educators that the root of the problem of underachievement for Hispanic and other minority students is quite often a matter of their low expectations for success in school. In a report compiled by John H. Rodriguez (1985), for the San Diego Office of Education, Gerald A. Rosander Superintendent of schools asserts that classroom teachers generally have very low academic expectations of Hispanic students. Moreover, he contends that it is not these expectations in and of themselves that are the malefactor, it is when such expectations inhibit teachers from being flexible. This failure to be open to new data, to provide students with challenging and appropriate instructional experiences, to give students deserved credit or praise, and to provide equitable opportunities, promotes a negative attitude toward learning on the part of the student. Furthermore, Rosander contends that these feelings of negativity propagate a non-caring attitude about the students' success in school and prevent effective teaching practices that might promote achievement success. Far

more serious, however, is the student's negative response to such teacher behaviors which reinforces the teacher's low expectations and causes them to be self-fulfilling prophecies.

Studies on Hispanic achievement have shown that in conventional classroom climates, teachers' expectations and attitudes of Hispanic students have been very negative; thus, the quality of instruction provided has been very poor. Consider, for example, the qualitative character of communicative participation and the opportunity to learn in the classroom. From the previous literature, it is inferred that conventional classroom settings prevent many Spanish speaking students from participation. Based on a major study of Mexican American education at the fourth, eighth, tenth and twelfth grades conducted by the U.S. Commission on Civil Rights (1973), Duran (1983) showed that the teacher student communication patterns were distinctly different for Mexican American as opposed to Anglo students. Teachers were found to direct praise or encouragement to Anglo students 36 % more often than to Mexican American students. In addition, teachers used or built on the spoken contributions of Anglo students 40 % more often than they did for Mexican American students; they also asked Anglo students 20 % more questions in class than they asked Mexican Americans.

Good and Brophy (1972) concluded that the clearest and most constant image a pupil receives in the classroom is the self image that is formed as a result of the teacher's reflections and interpretations of the pupil's performances and products. They further contend that the real experiences of pupils, successful and unsuccessful, are found in the nature of these reflections and interpretations and in pupils' responses to them. It is clear that in conventional group instruction environments, teachers do not provide optimal opportunities for optimal learning for all class members. If teachers have low expectations of Hispanic children to

begin with, then it is very unlikely that they will provide them with the type of attention they need for optimal learning to take place. Teachers are likely to demand better performance from those children for whom they have higher expectations and are more likely to praise such performance when it is elicited. Usually teachers give attention and positive reinforcement for learning only to high achieving students who comprise the upper one third or one fourth of the class (Anania, 1983). In contrast, they are more likely to accept poor performance from students for whom they hold low expectations and are less likely to praise good performance from these students when it occurs, even though it occurs less frequently. Furthermore, it has been shown that teachers with low expectations of their pupils can reduce the subject content and limit the activities, thus influencing their students' performance (Kasch and Borich, 1982).

In a classroom context, self perceptions of "competency" related to academic ability become the dominant manifestation of a sense of self-worth and thus a motive for maintaining credibility in the presence of the teacher as well as peers. Because ability is perceived to be a central ingredient to academic success, it is understandable that efforts to protect a sense of ability is a major preoccupation among students. Numerous studies show evidence that pupils of all ages, from kindergarten to college, value ability (Harari and Covington, 1981; Nicholls, 1975, 1976; Sohn, 1977) and particularly older students prefer to be seen by others as achieving by means of ability rather than by dint of personal effort.

The teacher's role in the classroom is extremely significant for the development of the student's sense of personal and academic worth. It is the teacher who ultimately has the responsibility for making final judgements of the student's performance. If the judgements are negative students may see the need to develop coping strategies to help deal with such negativity. Self-serving excuses can be one of these coping strategies. More specifically self-serving

excuses may be used to maintain an aura of public credibility. In the teacher-learning process students must balance the use of failure-avoiding strategies against the realities of teacher rewards and punishments. Many students thread their way between the threatening extremes of high effort, with its implications for low ability should they fail, and no effort at all. Effort in this sense is characterized as a "double edged sword" (Covington and Omelich 1979c). Excuses appear to be a basic ploy in achieving this precarious balance. It seems that the will to learn likely depends as much on the "certainty" with which the individual holds a given self-concept of ability and on the degree of discrepancy between publicly and privately held images, as it does on the level of self-perceived competency per se. It appears that as long as students are somewhat uncertain about the causes of their failures, even when such uncertainty is the product of defensive maneuvering, they may respond well to praise and success (Covington, 1984).

Teachers are not the only culprit in contributing to the development of low expectations in Hispanics. The institutions of which they are a part play a seminal role, too. Indeed the institutionally entrenched attitudes toward minority students is a social phenomenon that has been documented for many years. A Ford Foundation report compiled in 1984 includes a section on educational achievement with a study by Orfield (1980). Orfield argues that while researchers have had difficulty establishing the independent impact of schools on educational achievement, observers also believe that school quality, discrimination by school professionals against Hispanics, and teachers' self-fulfilling prophecies of low achievement for Hispanics have historically played a role in "pushing" Hispanics out of school. It is a historic fact, Orfield contends, that Mexican Americans in the Southwest have long been discriminated against in the public school systems. They have been segregated from non-Mexican-origin children and

punished for using Spanish even in casual conversation. They have attended schools that are poorly staffed and equipped. Orfield's study shows that while school segregation has generally decreased for black students in all areas of the country with the exception of the Northeast, the segregation of Hispanic students is on the rise from an already high level. The origins of this kind of disproportionate type of practice, which blatantly singles out Hispanic students, contributes to the prevailing negative attitudes held toward these students as well as the low expectation relative to their school performance.

Testing, Evaluations, and Hispanic Achievement

As Walker (1987) states "test bias with respect to Hispanic students is not a new area of exploration, but the nature of the tests, testing procedures, and interrelationships among tests are receiving more thorough scrutiny than ever before" (p. 29). As stated in previous sections, the use of standardized tests results to gauge Hispanic achievement has severely affected students' success since the results of such tests were used to place students in ability groups, and track them into remedial classes thereby perpetuating non achievement for these students. The continual use of results of standardized method to make instructional decisions has been a vital concern of education researchers.

Essentially the test content of standardized instruments does not reflect the classroom instructional objectives. The content is usually unimportant or irrelevant to what students need to know or understand. Furthermore, test content measures mainly recall-type learning, neglecting the higher thought processes--analyzing, synthesizing, and drawing generalizations and applying them to

new learning. Finally test vocabularies and illustrations are often unfamiliar to those who are not of white middle-class cultures or for whom English is a second language; most standardized tests are culturally and linguistically biased (Darehshori, 1977; Engle, 1977; McKenna, 1977; Quinto, 1977; Taylor, 1977).

Initially the scrutiny of standardized assessment measures to gauge achievement for educating Chicanos was done by George Sanchez (1932, 1934). Sanchez examined the Stanford Binet Achievement Tests and the Haggerty Intelligence Test and found that although there were marked correlations in all abilities tested, reading correlated most highly with all the other subtests. Specifically, he saw English language ability as one of the most important variables resulting in different IQ scores among the Spanish speaking subjects used for the study. He therefore concluded that the Binet tests of vocabulary for bilingual students were invalid as measures of intelligence. He indicated that the vocabulary used was inappropriate for these students. In addition he contended that simply translating a test from English into Spanish and expect it to retain its relevance for assesment of the intelligence of bilingual children is erroneous. He strongly urged an examination of the responsibility of schools toward bilingual children in the achievement of desirable educational goals. His position was that schools had the responsibility of creating relevant learning experiences for bilingual children.

Whereas testing has traditionally determined and reinforced perceptions of minority student inadequacy, several researchers have begun to question the efficacy of such testing procedures. More important, De Avila, Havassy, and Pascual-Leone (1976) have gone beyond Sanchez to suggest the kinds of tests needed for bilingual students and have examined the intellectual development of Spanish-speaking students by comparing the use of standardized measures and neo-Piagetian developmental measures. De Avila, Havassy,

Pascual-Leone (1976), and De Avila (1981) found no difference between Mexican-American and Anglo-American children on developmental measures but pointed out significant differences in their performance on traditional IQ and achievement measures.

Although it is beyond the scope of this review to treat all areas that have been considered in the field of testing, it is important to mention that achievement testing for Spanish speaking students has been researched in many ways. According to Duran (1986), assessment of language minority children's success in elementary and secondary school has received increasing attention over the past fifteen years. In a thorough review of testing of linguistic minority children, a section on the assessment of school achievement describes the various factors considered in the studies investigating achievement. All the studies included in this review deal with factors affecting achievement that are either linguistic variables or other variables like SES, ethnic affinity, aspiration levels, length of residence in U.S, or personal and background variables. One study examined the correlation of achievement test scores, aspiration levels and grade-point average, but not one study reviewed in this report dealt specifically with in-classroom practices (i.e. conventional group instruction) and how these practices affect the achievement of Hispanic students.

Conventional Group Instruction

It is clear from studies (e.g., Duran's 1983, 1986) that the conventional group instruction approach has not met the diverse educational needs of Hispanic students. Conventional group instruction does not provide optimal qualities of instruction for all members of a group. It is not designed to alter the entry characteristics of students in ways which could enable most to attain high levels of achievement and positive affect toward learning. Rather, it has been shown that the

amount and kind of reinforcement available to learners may depend on the level of achievement they attain.

It seems that Mexican American children start school fairly close to Anglos in all areas of measured achievement (Carter and Segura, 1979). The deficit is least with the young and increases with additional years of schooling. Competition may, in turn, result in tracking and ability grouping and, in its most blatant forms, racial/ethnic segregation. According to Carter and Segura (1979), rigid homogeneous ability grouping and tracking have several effects: one of the most important is isolation of one ethnic group from another. Low ability grouping is recognized as such by the students in them, as well as by those in high ability sections. Teachers and other school staff perceive and behave differently toward children in these groups. An ethnic group that predominates in low ability sections is stigmatized. Mexican American students have historically predominated the low ability sections of public school classrooms.

Mexican American students seem to eventually learn their assigned roles in classrooms. It is doubtful that these students develop the intrinsic motivation necessary for learning. Covington (1984), contends that "in order to provide sufficient rewards for developing intrinsic motivation for learning in the classroom, it is required that a basic alteration of competitively oriented achievement structures be made in the present system" (p. 97). In conventional competitive classroom environments there is little relationship between performance and reward because each pupil's rewards are contingent on the performance of others. Regardless of how well he learns, if others learn better, he will not receive the highest grades. Conversely, regardless of how poorly he might perform, if others perform less well, then he will receive the highest grades (Block, 1970).

Covington's thesis is that schools as institutions often fail.

Schools provide insufficient opportunities for success for students at all ability levels due to their competitive nature. Students are forced to compete among themselves for a fixed number of rewards (e.g. grades) that are unequally distributed within the classroom. Furthermore, this inequity is perpetuated by the practice of grading on a "curve" which allows a few students to succeed at the expense of many.

Studies by Ames and Ames (1981) and Felker (1976) conclude that not only are successes fewer in competitive environments but the very fact of competition alters the perceived reason for these successes (and failures as well) by exaggerating the importance of the role of ability in achievement. Such a perceived ability outcome linkage culminates in a state of demoralization for many students. As a consequence and as a way to protect their egos, many students start using strategies to avoid failure. The most direct way to avoid failure is simply not to participate. Students exhibit an "unwillingness to do work that is not absolutely required or do as little as possible on required assignments and, in its most extreme forms, absenteeism and chronic inattention are manifestations to avoid failure" (Covington, 1984 p. 83).

In addition, Covington (1984) states that "far more harmful are those failure-avoiding strategies required to meet the situation in which the student is obliged to participate, yet expects to fail" (p. 83). The majority of Hispanic students fall into this category. "Here, the student might not be able to avoid failure, but might at least attempt to side step the implication of failure. As a group these strategies seek to shift the personal causes of failure away from the internal attribution of ability toward external factors beyond the individual's control or responsibility" (p. 83). According to Weiner (1979), however, emotional reactions to external factors which are continually beyond the control of the student may eventually create intense hostility:

... it became evident that it is incorrect to presume an invariant positive relation between intentionality and the magnitude of emotional reactions in achievement settings. For example failure ascribed to others, such as the bias of a teacher or hindrance from students or family, will presumably generate greater anger and hostility. In this event externality is positively related to emotional intensity. (p. 61)

A prime example of this type of student is the "underachiever," who, by not trying, provides no evidence of actual ability and consequently experiences little shame when failure occurs. Indeed, underachievers often make a virtue out of inaction by downgrading the importance of the work they refuse to do. They may even come to take a perverse pride in failure as a mark of nonconformity (Bricklen and Bricklen, 1967).

Rodriquez (1985) concluded that although most school people want to be fair in their treatment of students, some school policies and practices nevertheless mitigate against equity, particularly in methods used to group students for instruction within the classroom and in tracking schemes. Both produce inequities in access to knowledge. The most blatant ways that conventional approaches to teaching have perpetuated inequities are through the competitive structure of school environments. These conventional classroom organizational structures which promote Hispanic non-achievement have been carried out through the practice of "tracking" and ability grouping. Sixty three percent of the Hispanic underachievers are Mexican Americans as reported by studies reviewed by Walker (1987). The

performance of these underachievers has been documented since 1930, but even today the achievement record is dismal especially at the secondary level (Walker, 1987). Tracking has played a significant role in the types of school programs offered and the resultant achievement experienced by Hispanic students. Tracking, Walker contends, most certainly eliminates opportunities and options for these students, reinforces minority student stereotypes, and reduces self expectations for performance. More directly, it affects general achievement and high school completions.

Carter and Segura (1979) further conclude that "such practices that isolate Mexican American children not only discourage equal-status interaction between them and their Anglo peers but also reinforce the stereotypes each group holds of the other" (p. 237). In addition, they contend that the track system strengthens teachers' stereotypes of Hispanics and thus affects their achievement. Since they have lower expectations of these students, the teachers make less effort, have lax achievement standards, and offer less encouragement. Thus Hispanics always end up at the lower end of the continuum of skills at any academic grade level.

Ogbu (1978) has presented specific evidence of how these classroom climates and assessment greatly influence Hispanic children's performance in school. In his Stockton, California, studies he found that "students who sense that their teachers think them capable of doing well and expect them to do well are usually successful, whereas children who believe the opposite tend to fail. This subsequent failure usually reinforces the teacher's expectations and, in turn, contributes to more failure" (p. 301).

One of the most significant ways student achievement is gauged is the way in which students' performance is assessed in conventional classrooms. In Ogbu's study, putting forth effort, did not seem to make a difference in the way students were evaluated.

Even when children exhibited more effort, they received the same grade of "C" year after academic year. Often children's efforts in the classroom are related to rewards in the forms of marks. Repeated "C" marks would hardly inspire Hispanic students to continue to put forth much effort.

Students' placement for learning on the basis of results of standardized testing is certainly an important aspect related to achievement. Conventional classrooms use normative standardized testing for evaluation of performance. Ogbu's study showed that as students were grouped and progressed through school, the groups become more and more defined. Ogbu (1978) contends that as a result of tracking or ability grouping the students in a primary school, for example, will be affected by the level at which they participate in a given course. Different reading and math levels cover different amounts and types of material. Placement in low levels may, therefore, dramatically limit students' participation. At the junior high, ability groups become separate classes and the teacher adapts the materials and techniques to the abilities of the students. Finally, in high school, the division into major curriculum areas in which students participate is primarily determined by the school's personnel (e. g., counselor, curriculum directors, or administrators).

As a reaction to traditional competitive classroom instruction which is based on norm referenced assessments and not designed to meet the individual needs of students, anthropologists such as Moll and Diaz (1987) examine how instructional contexts are socially constituted. These interactions they contend, are sociohistorical events that according to Scribner and Cole (1981) and Olsen (1986) play a critical role in learning. This study uses a Vygotskian approach for looking at the learning process. These investigators feel that for instruction to be effective it must lead students; it must be aimed not only at

weaknesses manifested in individual assessments, but at strengths that are displayed most readily in collaborative activities. Furthermore, the schooling of working class children they contend, has been conducted by the practice of reducing or "watering down" the curriculum to match perceived or identified weaknesses in the students. Thus, the problems working class children face in school should be viewed primarily as a consequence of institutional arrangements. They propose that certain institutional arrangements constrain children and teachers by not capitalizing fully on their talents, resources and skills.

One solution to these inequities proposed by these researchers is to utilize community-based knowledge texts and procedures to train teachers to use community resources. These investigators concluded that reading and communicative resources can be strategically combined or mixed to provide the children with the support necessary to participate and profit in reading and writing lessons in the classroom. Thus the successful learning for these children has to do with the social organization of instruction and how it interacts with children's and teachers' characteristics. The level of lessons, they feel, need not be reduced to accommodate children's English language constraints, but there are reasonable and credible ways to fruitfully relate lessons across languages for the benefit of the children. Presumably, children's Spanish language and literacy skills can be used to facilitate performance in English. These conclusions are based on the assumption that there is "no single developmental model in acquiring uses of language. And furthermore, just as academic failure is socially organized, academic success can be socially arranged." Thus, children differ in their abilities in ways that cannot be assessed solely by techniques that analyze independent performance. Finally, these

researchers present a bilingual corollary: children differ in their abilities in ways that cannot be assessed solely by techniques that analyze performance in one language.

Conventional Classrooms and Intrinsic Motivation

The literature on concepts of intrinsic motivation and achievement is directly related to the notion of competitive environments as conventional teaching approaches. This literature might explain what is occurring psychologically to Hispanic students that are labelled as low achievers. Intrinsically motivated behaviors are defined as those behaviors that are motivated by the underlying need for competence and self-determination (Deci and Ryan, 1980). The intrinsic motivation theory has evolved from the work of many researchers, (White 1959; Weiner, 1977; deCharms, 1968; Deci, 1975, 1980, 1981). The proponents of this view hold that all people are born with a need to feel that they are capable of making changes in their environment and that they can interact with it effectively. Thus, all students have basic needs for competence and self determination and seek challenges according to their individual capacities.

It appears that a lack of control over aversive events deters students from motivation for learning and cognitive development since students have difficulty learning from the effects of their responses. What seems to be essential for learning is effective action on the immediate environment resulting in intrinsic gratification. Moreover, attributions and their properties guide feelings. For instance, most students see ability as being a stable and uncontrollable factor, and effort as being variable and controllable, thus an attribution to effort is more motivating than

an attribution to ability. Ability can be seen as a factor which can be increased through effort and learning.

Deci (1975) points out that the need to have control is based in the central nervous system and is ever-present and motivates ongoing thoughts interrupted by basic drives or emotional responses. The need for competence and self-determination leads people to seek out and conquer challenges that are optimal for their capacities. When people are unable to behave in a competent self-determined way, they feel and display the response described by Seligman (1975) as "learned helplessness." He describes this state as a feeling of the lack of control over the outcome of aversive events. Hence, when placed in a nonresponsive environment (see Seligman's animal experiments, in Abramson, Seligman, and Teasdale 1978) one feels helpless and without a feeling of self-determination and needs to gain control. The experiments reported by Seligman (1975) conclude that a major consequence of experience with uncontrollable events is motivational. According to Seligman, it appears that uncontrollable events undermine the motivation to initiate voluntary responses that control other events. "A second major consequence is cognitive: once a man or an animal has had experience with uncontrollability, he has difficulty learning that his response has succeeded, even when it is actually successful. Uncontrollability distorts the perception of control" (Seligman, 1975, p. 30).

The core of self determination involves action on the environment to feel intrinsic gratification. Competitive classroom environments affect a student's self-determination. In order to feel self determining and intrinsically motivated, one must be effective in one's interaction with the environment and see results of actions (Deci, 1975). If students are not perceived by their teachers as competent, they begin to develop mechanisms to save

and protect their ego, and their performance and motivation are hindered.

Yet another line of evidence to substantiate the view that the competitive atmosphere of the conventional classroom results in diminished motivation for learning is Weiner (1974). Emotions according to Weiner are discriminably related to particular attributions. In this view attributions and their properties guide feelings. Weiner (1979) proposes that in achievement related contexts, there are multiple sources of affect following success and failure. These are ability, effort, task difficulty, and luck. The theory of attribution presented by Weiner suggests that ability is a stable uncontrollable factor and comparable to effort which is viewed as being variable and controllable. In addition, this view proposes that attribution to effort is more motivating than an attribution to ability. Student's motivation for learning decrease when they believe they cannot succeed.

Central to Weiner's thinking is that the motivation for learning depends highly on how students view ability. Several studies conducted by Nicholls (1978) suggest that the way that students perceive their ability depends in turn on whether they are "task involved" or "ego involved." The task-involved students view learning in a way that is similar to that of young children: the goal is to master the task, behavior is intrinsically motivated, and the reward is increased skill and competence. Ability is seen as a factor which can be increased through effort and learning. The ego-involved students' attitude is typical of older children: the goal is to prove to others and to oneself that one's performance is superior to others (or not inferior in comparison to peers). The task is only a means to this end, rather than a goal in itself. Behavior is ego involved therefore it is extrinsically motivated. Now if performance depends on how the student compares with other students, ultimately, the meaning assigned to the concept of

ability governs the self-concept, the will to succeed, persistence, and achievement efficacy. Bloom (1971), states that students continually judge themselves against standards set by themselves, teachers, and peers. Further he states that these relative judgements are made so frequently because the schools have for so long stressed competition as a primary motivational technique.

Summary

To summarize we have seen that conventional classroom environments are designed to meet the need of only a few students and promote competition which in turn creates divisions among groups. We have also seen that teachers develop low expectations of students who cannot compete and treat these students negatively and differentially. These systems use norm-referenced assessment procedures in order to make decisions about instruction, academic advancement, and placement. All these factors are interrelated and, taken together, are not responsive to the educational needs of Hispanic students, particularly the diverse groups that comprise Mexican Americans.

In conclusion, we reiterate that what is needed to off-set the consequences of conventional group instruction are interventions that will promote better and higher teacher expectations, non-competitive environments, and methods of evaluation that focus on what the students are actually learning rather than in comparing knowledge acquisition to an established norm that has nothing to do with what is being taught in the classroom. What is needed are equity structures that promote better attitudes toward all students, have a better way of assessing skill and promoting a better sense of academic self-esteem. And finally, what is needed is an equity

system that gets rid of competition and thus promotes intrinsic gratification for learning and deters the emphasis placed on an external reward system. Mastery Learning systems have shown positive results in all of these areas. Such systems may therefore be a viable means for instructing Hispanic students of diverse educational needs.

Mastery Learning Structures as an Alternative to Conventional Instruction

We can only determine the full limits of what the students can and will learn when we have provided qualities of instruction which are optimal for the individual learner Bloom (1976) . Hispanic students need equity structure classroom environments that allow high aspirations to be realized and also enhance the quality of educational experiences for all students.

Mastery Learning provides equitable opportunities since the teacher sets up the framework for instruction, learning, and evaluation in advance. The students are made to feel as if they will attain the goals established by the teacher and thus, will learn to trust the teacher (Spady, 1974). The instruction students receive is enhanced by providing feedback/correctives. Mastery Learning also features the use of absolute teacher-based performance standards such that any number of students can achieve a given grade so long as their accomplishments exceed the teacher's preannounced criteria.

Thus Mastery Learning is an equity structure that provides a basis for positive attitudes and higher expectations of student achievement, and eliminates competition for learning. Testing and evaluations are ongoing procedures directly related to instruction by

continuous feedback and correctives and, finally, tracking and ability grouping are eliminated.

The teacher's image of students in a Mastery Learning environment is positive since the approach to instruction will allow the teacher to take into account the learning needs of all students for optimum achievement. With an equity structure such as Mastery Learning, the teacher from the onset begins the academic year with an attitude which expects and demands better performance from all students, it is more likely that praise, encouragement, and positive reinforcement for learning will be more evenly distributed. In such an environment, the teacher's job is to develop in the student an interest in learning and a positive attitude toward school. If the student encounters success in learning, his self concept will be positive. Finally, the teacher can provide more reinforcement for learning in setting up the curriculum for a Mastery Learning approach. The learning and teaching tasks are delineated in advance and broken down into basic units thereby allowing the teacher to become more aware of whether or not the students are acquiring the necessary basic knowledge required to continue to advance from one level to another in a given subject.

The Mastery Learning approach to instruction develops intrinsic motivation for learning by setting standards for mastery and excellence set in terms of what is to be learned and apart from inter-student competition. Under Mastery Learning systems it is entirely conceivable that all or none of the students may attain mastery. Each student is appraised individually solely with respect to his performance vis-à-vis a fixed standard rather than his performance relative to a group of his peers. By setting realistic performance standards developed for each school or group, followed by instructional procedures which will enable the majority of students to attain them, an equity structure for instruction is

established which sets achievement standards that place emphasis on cooperation rather than competitive learning (Bloom, 1970). Such an approach might eliminate the inequities that exists in a competitive system that has failed to meet the needs of Hispanic students. An equity system like Mastery Learning will shift the task emphasis from ability to effort, thus produceing a positive effort linkage which may culminate in higher success in achievement for Mexican American/Chicano students.

A Mastery Learning approach features the absolute teacher based performance standards such that any of the students can achieve a given grade so long as their accomplishments exceed the teacher's pre-announced criteria (Block, 1977). Equity learning structures like Mastery Learning emphasize achievement via effort and discourage low effort in order to strengthen a positive affect linkage. Bloom (1978) proposes that Mastery Learning is one of several strategies that can succeed in bringing a large proportion of students to a high level of achievement and high motivation for further learning. Additionally, favorable learning conditions have profound effects on student learning attitudes and interests.

Central to most Mastery Learning strategies is the use of feedback and corrective procedures at various stages or parts of the learning process. While a variety of feedback processes are possible it has been found that the development and use of brief diagnostic tests are most effective. The effectiveness of Mastery Learning work is clearly related to the degree of efficiency of Formative Tests in pinpointing the learning needs of each student. The key to success of Mastery Learning strategies largely lies in the extent to which students can be motivated and helped to correct their learning difficulties at the appropriate points in the learning process (Bloom, 1978).

Reports by Block and Burns (1977), present evidence that students receiving feedback/correctives exhibit higher levels of

achievement and positive affect than found in conventional instruction. Mastery Learning strategies have been shown to enable students both to attain realistic levels of skill that may not have previously been possible in other contexts and to obtain recognition for their accomplishments through certification mechanisms that reflect actual rather than relative performance (Spady, 1974). This in turn, it seems influences the affect the students develop toward learning in general (Block, 1970; Nordin, 1979). Recently, Block, in Block, Efthim, and Burns (1989) contends that "Mastery Learning procedures should have the capacity to make ordinary school "work" activities seem like "play." (p.54).

In addition to the obvious function of the feedback corrective procedures (that of enabling students to reach high levels of learning before proceeding to new tasks), it is likely that what occurs during the procedure serves to increase student motivation for further engagement in learning. The Feedback and Corrective Loop enables almost all to obtain evidence that they learned well and are capable of successfully meeting the demands of the task. The relation between affect and perception of achievement in Aniana's (1984) study is particularly strong. This study showed that when students received a quality of instruction adapted to individual learning needs, prior achievement and aptitude had little influence on the achievement they were able to attain. Studies such as Levin's (1979) and Nordin's (1979) have shown that providing feedback/correctives resulted in significantly higher levels of summative achievement, total time on task, and positive affect than occurs for students receiving conventional instruction. It appears that the role of students' characteristics in determining learning outcomes depends on the quality of experiences provided.

A very important dimension of Mastery Learning structures is that, unlike conventional classroom instruction, tracking and fixed ability grouping is eliminated. According to Block et al. (1989) equity

structures like Mastery Learning offer opportunities for all since there is no homogeneous grouping to ensure that some students master different products and processes. In addition, these structures are characterized as "equalence" structures and therefore can use existing resources to provide economy in teaching. These approaches to teaching employ evolutionary rather than revolutionary methods for meeting educational needs such that the focus is on altering things that can be altered. Furthermore, the focus on student learning is on prevention rather than remediation.

Block (1974) reports that at least four experimental studies (Austin and Gilbert, undated; Block, 1970 or 1972; Capongri, 1972; Lierly, 1973), and some non-experimental (King, 1971; Shepler, 1969) show that mastery approaches to the teaching and learning process may help to off set the negative effects of low ability on student learning. Most of these studies suggest that improved achievement of lower ability students under Mastery Learning conditions does not seem to come at the expense of achievement of the higher ability students. Individual differences in entry characteristics would not yield individual differences in their achievement. Each student would be helped by a feedback/corrective system to achieve the initial learning units to the same level regardless of their entry characteristics.

More recently, Block et al. (1989) proposed that under a Mastery Learning approach students' learning includes not only certain intellectual outcomes but also certain emotional and behavioral ones. Block contends that what is most important is learning-to-learn skills that assume self-care for learning and personal responsibility. Moreover, he proposes that under a Mastery Learning paradigm students will develop "response-ability" for learning what they emotionally can and want to do. This self treatment of learning or self evaluation coincides with Bloom's (1970) early notion of self-assessment. Assessment, along with

measurement and evaluation, is one of three major schools of testing where self-examination results in the expectation that a better alignment of actions with thoughts and feelings results in mutual respect between teacher and student.

Ultimately by using Mastery Learning as an equity structure the most important outcome is higher achievement for the majority of students. Guskey and Gates (1986) provide a synthesis of Mastery Learning research studies that examine the effects of group based Mastery Learning programs. In this analysis student achievement was the primary variable. These studies show that achievement results are overwhelmingly positive. Of particular interest for the present study are the results that pertain to student learning strategies. It seems that students are engaged in learning for a larger portion of the time they spend in mastery classes and require decreasing amounts of remedial (corrective) time over a series of instructional units. In addition, students in Mastery classes seem to develop positive attitudes about learning and about their ability to learn.

Summary

To conclude, we have seen that conventional classroom instruction approaches propagate an overwhelming number of negative attitudes and expectations because of the competitive nature of these structure which are based on norm referenced assessment. These practices result in low achievement and low motivation for learning for the majority of students. On the other hand, we have seen that there is evidence that Mastery Learning classroom approaches offset the negative effects of competition since students are given an opportunity to mastery the basic material they need to learn for further learning. Assessment is done objectively since teachers use criterion-referenced diagnostic procedures in order to

periodically gauge achievement and the on-going results are used to provide continual feedback and correctives. Finally, all students are expected to learn, not just a few, regardless of prior achievement. And most students will be intrinsically motivated to learn since they will feel a sense of success in learning when they master the learning task.

There is little direct evidence however that Mastery Learning structures are successful in situations where the majority of students are Mexican American/Chicanos. Block et al. (1989) report results of a study which included a student population that was 65% Mexican-American background. The results of this study following a three year implementation of Mastery Learning show that the average level of student achievement on the California Test of Basic Skills in reading, language, and math was higher and also that student achievement was homogenized around these high levels. In other words, during the three year period (1983-1986) three out of every four students were above the 50th percentile in all three content areas.

Block has also been working closely with a local elementary school that has an enrollment of 70% Spanish surname students. This particular elementary school has implemented Mastery Learning in all content areas for the last five academic years. An examination of the standardized tests scores on the California Achievement Tests in reading, language, and math show a continual increase in achievement. The results of the scaled scores (stanine scores) from 1984 to 1987 for the third grade indicate a 44 point increase in reading, 39 point increase in language and a 15 point increase in math. For the sixth grade there is a 42 point increase in reading and quite a bit higher for language and math; 60 and 82 points respectively. During a phone conversation with the school's principal, she explained that most of the teachers implement a Mastery Learning approach and all of the instructional

objectives are aligned with the district's instructional plan. In addition, teachers organize half hour to two hour sessions for practice and correctives after school at least 2 to 3 times per week. Indeed, the achievement results of this school have been so positive that it has been selected as an exemplary model school for the State of California.

An independent school district located in south Texas has also recently implemented a district wide Mastery Learning program. This district is comprised of 85% Mexican American students. During phone conversations (Fall 1989) with the district superintendent, program evaluator, and resource teachers, it was stated that the basic emphasis of the program was to increase achievement in all basic skills areas, but that there was a special focus on improving the basic writing skills of students. The assessment procedures that are being utilized are many since the approach for both teaching and assessment focus on process as well as overall outcome. Unfortunately, since this program has just recently been implemented, no documentation of achievement results was available for examination. It was stated however that the overall attitude of the teachers, who are striving for excellence in learning for all students, has improved 100%.

Since there is not enough available evidence to state with confidence that Mastery Learning programs work better than other methods to raise achievement and educational aspirations for the majority of Mexican Americans, more studies like the present one are needed. The premise and content of the present study is that in order for students to succeed in an academic setting they must develop fundamental writing skills. This requires that they develop ways of logically organizing ideas for clarity of expression. In addition they need to refine conventional usage of grammar and sentence and paragraph structure. These are essential skills for academic success at any grade level beyond the third grade. The

basic writer is unsure and lacks confidence in his or her ability to express ideas in writing. Writing is an expressive skill that is ego involved; it reveals the self and far too often students hesitate to express themselves for fear of rejection and criticism from the teacher and other students. Very few Hispanic students are successful the first time they are taught how to write. The classroom social climate (from personal observation) prevents many Hispanics from taking an active role in pre-writing activities. Consequently, most of these students do not develop adequate expressive skills (as judged by subjective teacher norms) and usually are evaluated and placed at the lower end of the spectrum in basic skill level. These students are alienated and are usually perceived as low-ability students.

From all the evidence presented by previous Mastery Learning studies it seems reasonable to propose that a Mastery Learning approach for teaching Hispanic students basic writing skills is a better approach than a conventional approach. This will raise the achievement level of Hispanics. High achievement will result in good perceptions of ability and successful, academically capable individuals will thus have positive attitudes and interest toward learning.

The study proposed to focus on the relation between achievement and the instructional approach used. Basically it examines and compares two learning conditions, conventional group instruction and Mastery Learning. Because of contextual conditions it was necessary to conduct this study in two phases. The original plan was to compare two learning conditions; this was done in Phase I of the study. But as we shall see, problems were encountered in implementing the Mastery Learning component of Phase I (referred to as the original study), therefore it was necessary to implement Phase II (referred to as the revised study) where only a Mastery Learning approach was utilized (see Chapter III and IV for

elaborated explanations). In both phases of this study however the primary question was: Is the outcome of achievement a function of the instructional approach used? In addition, the study focused on the relation between achievement and two dimensions of affect: interest and attitude toward learning. The following question was addressed: Do attitude and interest toward learning develop as a function of the achievement students attain and as a result of their perception of the adequacy of their achievement?

CHAPTER III

Methodology: Original Study Phase I

The Research Model

The study was designed as an exploration of the effectiveness of a Mastery Learning program to teach basic writing skills to Spanish language background students. To conduct the exploration, this study initially adopted elements of a research model developed and validated by Anania (1983) that compared the learning effectiveness of three different instructional procedures: conventional instruction, Mastery Learning, and individual tutoring. These three approaches were ranked at three levels from maximal to minimal with regard to quality of instruction: tutoring as maximal, conventional instruction as minimal, and Mastery Learning in the middle rank of quality of teaching.

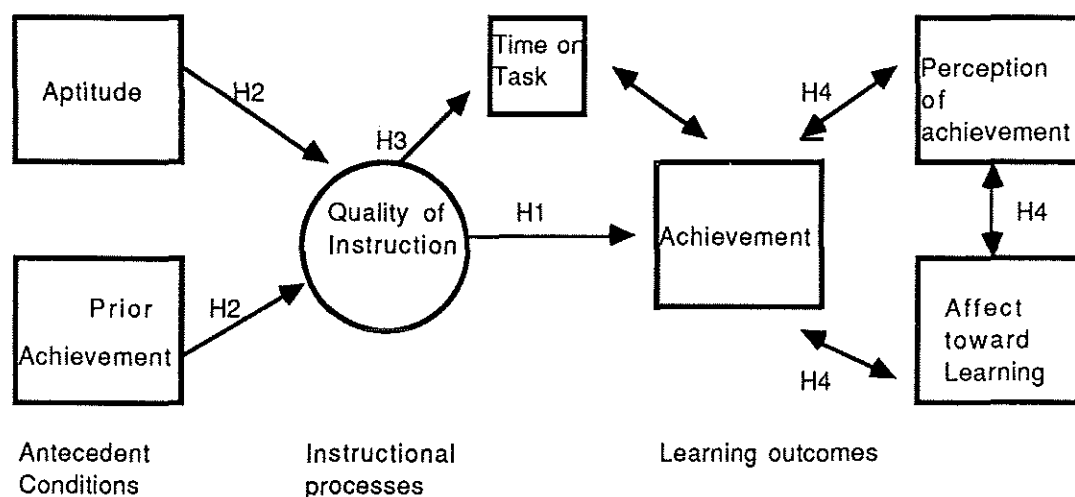


FIGURE 3.1 Model of the Effects of Quality of Instruction (Anania 1983)

Figure 3.1 illustrates Anania's model. As is evident in the model quality of instruction is the prime variable. The model posits that students' time on task and achievement are interactional and both depend on the quality of instruction. In addition, the model proposes that achievement, perception of achievement, and attitude and interest interact too.

The present study utilizes the three basic components of Anania's model: antecedent conditions, instructional approach, and learning outcomes. The Antecedent Conditions were operationalized as follows: Quality of Instruction was translated into two forms of whole group teaching, either a conventional approach to teaching or a Mastery approach; no tutorial approach was used.

Because of limited resources, the present study did not deal with the Time on Task variable as extensively as Anania's study. In Anania's theoretical model the Time on Task variable was a main component of the interaction between the quality of instruction and achievement. The variable Time on Task in the present study appears as an inherent component of the Mastery Learning Treatment classroom. In other words, it is not dealt with as a dependent variable per se, but is contained in the Feedback and Corrective Loop of the Mastery Learning approach. Time on Task in the present study is observed when corrective instruction is provided to the student from the results of various alternative Formative Tests and as the motivation for continual learning increases as a result of continual feedback. This Time on Task dimension however is not considered in the conventional approach used in the control classroom of this study. Some important additional variables in the original design were attitude, interest, and the student's perception of his or her own achievement. In sum, the variables for the study appear as Figure 3.2:

Figure 3.2 Schema for the Present Study Phase I

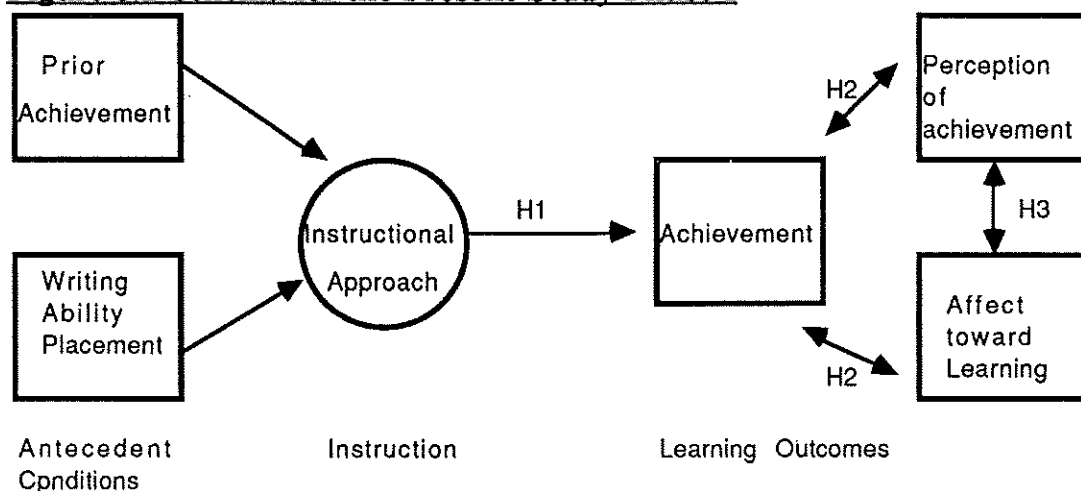


Figure 3.2 Schema for the Present Study

Research Hypotheses

Anania's model (see Fig. 3.1) hypothesized (H1) level and variation in student achievement to be a function of the quality of instruction given to students. This model also posits (H2) quality of instruction as intervening between effects of students' prior characteristics (achievement and aptitude) and their subsequent achievement. The model further hypothesizes (H3) that the level and variation in the percentage of time students are actively engaged in learning is a function of the quality of instruction they are given. Finally, the model proposes (H4) that attitude and interest toward learning develops as a function of the achievement students attain and of their perception of the adequacy of achievement.

The present study has three hypotheses as opposed to Anania's four hypotheses model (see Fig. 3.2). The first hypothesized (H1) that level and variation in student achievement is a function of the

instructional approach². It was expected that the highest levels of achievement and the smallest variation would be found in the Mastery Learning group rather than the conventional instruction. Mastery Learning incorporates procedures for enabling students to enter new learning tasks with high cognitive and affective entry behaviors.

The second hypothesis addresses the question "do attitude and interest toward learning develop as a function of the achievement students attain and of their perception of the adequacy of their achievement?" It was hypothesized (H2) that the combination of high achievement and positive perception of achievement should affect high levels of positive attitude and interest toward learning.

The third hypothesis (H3) posits that level and variation in the active engagement of learning is a function of the quality of instruction. Because Mastery Learning provides a more favorable quality of instruction than conventional instruction the attitude and interest in the Mastery Learning groups is expected to be higher than found for the conventional group.

Hypothetically, attitude, interest, and perception of achievement are tightly dependent on the instructional approach. The attitude and interest expressed by students taught with the conventional approach was expected to involve large disparities. It was assumed that positive affect and academic success is accessible to only a minority under conventional instruction conditions of learning. Since the conventional approach does not incorporate a procedure which enables students to

² Note that in the present study the instructional approach is the primary independent variable. To reiterate, the instructional approach was either mastery learning or a conventional approach. This is considerably different than Anania's (1983) study. Anania's quality of instruction included a conventional approach, a mastery learning approach and individual tutoring to compare the degree of interaction with other factors. The present study did not include individual tutoring nor did it consider the percentage of Time on Task as a measurable variable. Although Anania's model was used to generate a theoretical model, the schema for the present study (Figure 3.2) illustrates a sufficiently different model.

acquire the prerequisite learning they need to succeed, students will find themselves increasingly unable to comprehend the instruction. What Nicholls (1984) describes as a shift from task involvement to ego involvement can clearly be observed as a negative by-product of this type of instructional setting.

The Research Design Phase I

The design of Phase I of the present study conforms to quasi-experimental Design 10 identified by Stanley and Campbell (see Figure 3.3). This design involves an experimental group and a control group which are both given a pretest and a posttest but in which the control group and the experimental group do not have pre-experimental sampling equivalence. Rather, the groups constitute naturally assembled collectives such as classrooms, as similar as availability permits, but yet not so similar that one can dispense with the pretest (Stanley and Campbell, 1963). Since both teachers were using the same textbook prior to the onset of the experiment, effort was made to insure that the teachers were teaching the same unit at the same time. This procedure further approximated the groups and since these classrooms are similar in recruitment, the main effects of history, maturation, testing, and instrumentation were better controlled. This selection approach, therefore, provided better internal validity for Phase I of the experiment.

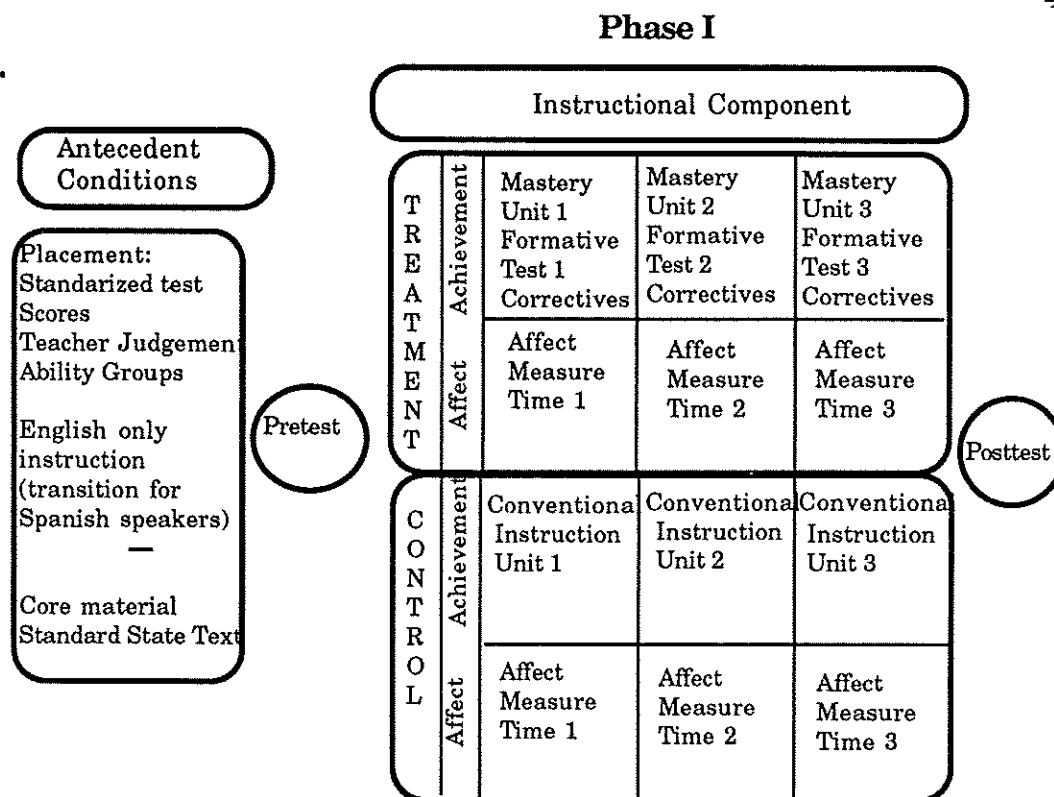


Figure 3.3: Pre Post Quasi Experimental Non-Equivalent Control Group Design 10 Stanley and Campbell (1963)

Developing the Dependent Variable

Achievement, the primary dependent variable, was measured in two distinct ways. The initial gauging was done at the beginning of the course by the use of a Pretest and at the completion of the four unit instructional sequence by a Posttest. In addition, Achievement was also gauged at the completion of each instructional unit by Formative Tests and, after the use of correctives at the conclusion of two instructional units, by a Summative Test (See Appendix A for samples).

Motivation as a dependent variable was gauged by the use of a three part questionnaire. This questionnaire concerned students'

perception of achievement, attitude, and interest in writing and was administered and analyzed each week in both control and treatment classrooms.

Gauging the First Dimension of the Dependent Variable: Academic Achievement CSE Subscales and Trait Scales

The Measurement Scales CSE and Primary/Secondary

The dependent variable Academic Achievement was assessed from two different perspectives. The first dimension was the Pre and Posttests writing samples which were evaluated by six distinct measures: four UCLA Center for the Study of Evaluation (CSE) Expository Subscales and a Primary/Secondary Trait (rubric) developed for criterion-referenced writing tasks. The CSE scales are based on generally recognized features of good writing, regardless of the topic or intended audience of the assigned writing task. The CSE Subscales used for this study allow a cross comparison of the performances of the students and classes. It is a widely used method for evaluation of writing, for example, the analysis of text elements. The Subscales also allow a comparative approach (with primary traits) of evaluating writing by different definitions of the writing skills construct. The Primary/Secondary Trait rubric, on the other hand, assumes that writing performance is highly affected by context, content, topic, purpose, and intended audience. Unlike the analytic and holistic categories of the CSE Subscales which can be used without modification for a variety of expository tasks, the Primary Trait is a task-specific, "holistic" scoring method. Its scoring rubrics are built from careful analysis of the features of the proposed writing task (Lloyd-Jones, 1977).

Any concept or idea produced by the student in either Spanish or English was acceptable as a writing sample and deemed suitable for the purposes of this study. The writing assignment used to generate the Pre

and Post writing samples for Phase I was determined by several factors: the concreteness of the topic, familiarity of the themes and ideas to the entire group, maturity level, knowledge of the concept, and socio-economic levels. The researcher and the teacher for the treatment group decided that since holidays is a theme that is used for instructional purposes and one that most, if not all, students directly experience, it would motivate the students to write.

The directions of the Pre and Posttests asked the student to write a paragraph about what a particular holiday meant to them and their friends and family. Moreover, the instructions tried to elicit a response by asking them to describe activities or experiences dealing with the holiday in question (either Halloween or Thanksgiving). For the benefit of the students who may have only recently been exposed to these holidays, the directions for the Posttest holiday theme (Thanksgiving) asked them to write about what they have learned at school about the holiday's meaning.

We assumed that most fourth grade students would be exposed to Halloween activities both in and outside the home. Even though holiday topics are dealt with extensively in most U.S. elementary school classrooms they may be somewhat exclusive of the newly arrived immigrant students. For instance, Thanksgiving is not experienced in the homes of many immigrant students. Therefore, the student had to rely on what was introduced and taught in the classroom. It was decided, however, that since all of the newly arrived immigrant students participating in this study were from Mexico, they had been exposed to these holidays either through the media or through their close geographical proximity to the U.S. It was assumed that Mexican children know, at least at a basic level, the significance of these U.S. holidays.

Gauging the Second Dimension of the Dependent Variable: Academic Achievement

Formative and Summative Tests

Unit Formative Tests are tests that are intended to provide teachers with information on how well students have mastered the unit goals, what goals, objectives, or which content elements are causing difficulty for the groups of students or individual students. They can be properly called diagnostic progress tests (Block et al. 1989). Their ultimate goal is to recognize degrees of competency and proficiency as they pertain to students' accomplishment of these specific learning outcomes.

For this study the specified learning outcome is proficiency in writing a basic paragraph which are components of a unified thought and which also require distinct levels of conceptualizing. Formative Tests were based on the content and objectives of the learning tasks. In Phase I these objectives were identical to those of the adopted state textbook. The tests served as a measure of diagnostic progress in the Mastery Learning classrooms and indicated where correctives were needed for students that did not meet the 85 percent criterion level. Similarly, the students in the conventional group were assessed by the same tests as the Mastery Learning group but received information about their scores only from weekly quizzes; they were provided no correction on the basis of these scores.

The unit Summative Test used in Phase I, was comprised of the instructional objectives which were included on the Formative Tests (and the results of the Formative Tests that were administered.) The items included the most important goals up to that point of the instructional sequence. The Summative, like the Formative Tests, were developed from the objectives stated in the "Table of Specifications", but

the important distinction is that the Summative Test constitutes a cumulative assessment of all instructional units.

Gauging the Third Dimension of the Dependent Variable:

Motivation

Perception of Achievement, Attitude and Interest

Motivation as a dependent variable in this study is operationalized as an interactional process which occurs as the result of a teaching approach that either promotes a good perception of achievement or raises academic self-esteem. Thus a student's perception of achievement was gauged by questions that dealt with how they perceived themselves with respect to their peers, their teacher, and the academic tasks. In addition, their attitude toward the subject and interest in learning the subject are components that were included under the definition of Motivation.

Developing the Instrumentation

Phase I Components

CSE Scale for Evaluating Writing

The CSE Expository Scale IV was originally devised at the center for the Study of Evaluation, UCLA, as one of a series of rating scales for scoring expository essays. The scale includes a combination of analytic and holistic rating categories and includes six subscales. The first two

subscales, General Impression and General Competence, involve holistic evaluation of the overall quality of the essay. Four subsequent subscales, Essay Coherence, Paragraph Coherence, Support, and Mechanics, provide for an analytic evaluation of these features in a given essay.

For purposes of this study and since the writing task requires students to produce a single paragraph rather than an essay, only four subscales were used: General Impression, General Competence, Paragraph Coherence, and Mechanics. In discussing the propriety of using these scales for evaluating the writing samples, an expert on research in composition, Dr. Sheridan Blau, concurred that the elements of a paragraph are structured characteristics of a broader unit such as an essay. Therefore, with minor modifications of the wording of the scale categories, they served as an objective tool for the evaluation of single paragraph writing samples (see Appendix C for modified descriptions).

The General Impression subscale (CSE 1) allows the raters of the writing samples to determine subjectively whether the writing sample is in general well written. The raters include their subjective reactions to originality of ideas and personal style.

The General Competence subscale (CSE 2), on the other hand, requires the rater to assign an overall quality rating as a function of competence in each analytic subscale used: elements of paragraph coherence such as indentation, sufficient number of sentences to express a unified thought or main idea, and mechanics such as sentence construction, usage, spelling, punctuation, and capitalization. The remaining two subscales, Paragraph Coherence (CSE 4) and Mechanics (CSE 6), asks the raters to "reread" the writing sample for each analytic characteristic and make a discrete judgment of quality for each element.

The subscales were used in the order presented; thus the raters worked from larger elements such as the unity of ideas within a paragraph to the elemental unit of the word itself.

Primary/Secondary Trait for Evaluating Writing

Primary Trait scoring is a system of scoring that is task-specific. The writing assignment is first analyzed to identify the major features of writing that should be included in ideal sample responses. Then, these features are quantified into a scoring scale; writing sample raters are trained to evaluate the utilization of these "primary" traits relative to the writing sample as a whole or to note their absence (Lloyd-Jones, 1977). The Primary Trait method of evaluation has been highly criticized for its one-task specificity which excludes other indications of a writer's competence beyond the specific task at hand. For the purposes of examining students' responses on a single occasion however, it can provide a reliable, holistic evaluation.

For purposes of this study, the Primary Trait method was used to develop primary and secondary scoring rubrics specifically by utilizing the pre-identified criteria from the Mastery Learning specifications used for instruction with the treatment groups. The Secondary Trait scoring rubric was developed to allow creative language to be considered in the overall evaluation since there appeared to be a sufficient number of writing samples that were "creative in nature".

The two Trait Scoring rubrics were developed in conjunction with rater training. Three paid raters and the researcher participated in the development of the Primary and Secondary Trait rating rubrics. The three raters, two Spanish/English bilingual graduate students, one English monolingual experienced writing teacher, and the researcher participated in the rating of writing samples with the CSE and Primary Trait scales. Since the Primary/Secondary Trait scales were developed

in conjunction with rater training, these rubrics were applied shortly after the CSE Subscale ratings were completed.

Formative and Summative Measures

Since the study collected data from Formative and Summative Tests administered three times during each phase of the study, items for the Formative and Summative Measures were criterion-referenced. They were developed based on the instructional objectives for the Mastery Learning classroom. During Phase I, the Formative Tests were used to gauge the achievement of the subjects in both the Mastery Learning class and the control class. Each of the three instruments used in both phases were customized to conform to the elements of the classroom instruction each week. Consequently, they vary in length, scope, and focus.

At the onset of the study, the researcher met with the teachers to identify, sequence, and organize the performance objectives in accordance with Bloom's Taxonomy (1956) for four consecutive weeks of instruction. After the objectives were specified, the material was reviewed and the Formative and Summative Tests were developed based on how the concepts for instruction were going to be presented. This process was used for both phases of the study.

For Phase I of the study, the control group teacher was given a list of the basic concepts to be covered during the four week instructional sequence and a specified list of pages corresponding to the state-adopted text that dealt with the relevant instructional concepts. These text presentations were used in the experimental classroom for corrective or supplemental materials and also to develop the content of some sections of the Formative Tests and the Summative Test. Both groups in Phase I were tested with the same instruments.

Formative Test I Part A and B (Phase I)

Formative Test I was developed in accordance with Part 1.1 of the Table of Specifications (see Appendix A). The content of the items in both parts A and B dealt specifically with word meaning and were related to the objectives of the lesson presented during an instructional sequence. The pre-determined skill level, assessed by the test, reached a level of application in Bloom's Taxonomy. In other words, the students were to know and comprehend the meaning of the words and apply them.

In both sections of the test, students were asked to circle the correct response. There were twenty items in Section A which assessed word and sentence comprehension. Out of twenty items, 12 were correct responses, and were eight, distractor items. Only the correct responses were counted as final responses for analysis. For instance, if a student had eight wrong responses and four correct responses only four were counted as actual responses. Similarly, in Part B, there were a total of ten items, but only seven items were correct responses (three items were distractors); the correct responses were counted for analysis. The following are sample items from Part B of Formative Test I which assesses sentence comprehension:

Sample item Formative I Part B:

Circle each group of words below which tell you something clearly:

1. I think does nice.
2. It broken came loud.
3. Joe is from another neighborhood.
4. The banana smells delicious.

=====

As can be seen from this Sample Formative Test I, Part B merely assesses word and sentence recognition skills.

Formative Test II Parts A, B, C (Phase I)

Formative Test II Parts A, B, and C assess the Content Objectives for Part 1.2 and 1.3 of the Table of Specifications. The items in each section of the test are designed to coincide with the instructional sequence and the primary objective is to test capitalization and punctuation. Four out of eight content objectives included in these sections reach a level of analysis in Bloom's Taxonomy for student learning behavior (see Appendix A). This means that the pre-determined level of learning that is required includes knowledge, comprehension, application, and analysis.

There were six items in Section A; students were asked to circle the appropriate punctuation mark. Part B is composed of four items but required eight responses, two for each item. In that section, students were required to utilize capitalization and punctuation. Part C items ask the student to re-write two sentences incorporating correct grammar and punctuation. In order to do this correctly, it required four responses. The following Sample Item assesses sentence capitalization and punctuation in Part C Formative Test II:

Sample Item Formative Test II Part C:

C. Re-write the following sentences by using the correct capitalization and punctuation:

1. i love baked apples
2. does Joe have a dog

=====

This test attempts to synthesize skills of word and sentence comprehension and mechanical skills. The focus however is on mechanics.

Summative Test (Phase I)

The Summative Test was administered at approximately the fourth week of the instructional sequence. The objectives covered

Sections 1.1, 1.2, and 1.3 of the Table of Specifications, the same skills covered by the two Formative Tests. To reiterate, this test unlike the Formative ones, is cumulative in nature. Essentially the test required the student to group thematically related sentences and to punctuate and capitalize the appropriate elements. The items were composed of five responses that could be ordered sequentially by theme (and two distractor items). Only correct responses were counted for analysis. For instance, if students used the distractor items or did not sequentially order the sentences, they were not given credit. The Phase I Summative Test holistic sample item is the following:

Sample item Summative Test (Phase I)

A. Choose the sentences that are about the same idea and group them together in order. Remember to use right capitalization and punctuation:

- 1.the teacher gave me a good grade on my homework
2. yesterday I had some homework
3. i rode the bus to school
4. my mother helped me do my school work
- 5.it took me half hour to do my homework
6. i felt good in class when the teacher checked my homework
7. the weather is getting cold

=====

This Summative Test III, as well as Formative Test II, focuses on skills of sentence mechanics. Summative Test III, however, attempts to assess idea sequencing skills as well as word and sentence comprehension. This Summative Test III was used to assess overall achievement of three instructional units.

Questionnaire to Test Motivation

Data was collected by the use of a three part questionnaire. This questionnaire, administered three times during a four week period in both phases of the study, was used to gauge overall motivation. The

responses to this questionnaire were collected at the end of each unit of instruction and were a composite of three areas of affect: Perception of Achievement (Part A), Attitude Toward Learning to Write (Part B), and Interest in Learning to Write (Part C).

The present study included a motivation component which was used to investigate and explore the relationship between achievement and self-concept of ability. It was initially assumed that if students developed a sense of success by the Feedback and Corrective Loop in the Mastery Learning classes, then they would develop a good self-concept of ability, thus be motivated to continue learning and ultimately achieve. Since the focus of this study is the interaction that occurs within the classroom, the teacher was viewed as the significant other. Since this study focuses on classroom interactions and the connection between achievement, teaching approach, and student's self concept of ability are central issues, the questions on the questionnaire were designed to focus on different aspects of the content of the Mastery Learning instructional sequence.

Items for the Attitude questionnaire were adapted from questions used in Anania's study which were originally developed and adapted from the Brookover Self-Concept of Ability Measure (1962) and from scales developed by Dolan (1974). After developing the content and format of the questionnaire in English, it was translated into Spanish. The initial translation was done by the researcher. Then both the English and Spanish versions were reviewed by a bilingual fourth-grade teacher, who was not involved with the study, to determine if they were appropriate for fourth graders. This teacher then presented several sample items to students and made a few suggestions for word and format changes.

Developing The Independent Variable

The independent variable in Phase I of the study was the Instructional Approach. There was an attempt to isolate the variables in the model in order to gain a better understanding of the relationship between them. The two instructional approaches in Phase I are Conventional Group Based Instruction and Mastery Learning.

The Conventional Instruction

The Conventional Approach classroom served as the control group in Phase I. This type of instruction is group-based and is generally directed toward the students who are most able to benefit from the instruction: the high achieving students. Students are usually tested at the completion of assigned tasks which provide the basis for evaluation in the form of final grades. They proceed to new learning tasks regardless of their performance on previous tasks.

With a Conventional Approach to teaching the levels of achievement can be predicted on the basis of aptitude and prior achievement. There is no method for correcting errors in learning or of assuring that most students acquire the cognitive entry behaviors they need to benefit from instruction (Anania, 1983). Under this instructional approach, perceptions of inadequacy lead to apathetic or negative attitudes toward the task and learning in general. The constructs described by Covington (1984) in his "self-worth" theory are manifested in this type of setting and result in low achievement and the development of a low self-concept. This phenomenon is assumed to develop when students are continually compared with their high achieving peers.

The Mastery Learning Approach

The Mastery Learning approach was the independent variable to be tested. This instructional mode was also group-based, but feedback from diagnostic, Formative Tests provided information for individualizing instruction. An 85 % criterion level was set for mastery of a learning unit. Students who did not meet the set criterion on a Formative Test after completion of a task were given additional opportunities for learning the same concept through the use of correctives. They were not supposed to progress to the next level of learning until they acquired the prerequisite knowledge to succeed. Correctives also served as enrichment activities for students who were able to move ahead at a faster pace. The correctives and feedback materials were varied and allowed additional reinforcement and individualization of instruction.

With a Mastery Approach, a feedback/corrective process enables the student to acquire the cognitive entry behaviors needed to succeed at new learning tasks, despite prior characteristics. Through continual feedback and correction the student views the effort put forth as leading toward success. Students enter each learning task with an optimal readiness for achievement. This base provides them with a better chance for developing a higher academic self concept. With a system that utilizes an instructional approach where feedback is given continually and instruction is periodically individualized through the use of correctives, it was expected that at least 85 % of the students in a given group will perceive themselves as successful and as academically capable individuals.

Figure 3.4 illustrates a few of the basic differences between Mastery Learning instruction and a Conventional Approach. These areas of comparison were selected because they seem to be the most pertinent for this study.

Figure 3.4: Comparison of Independent Variables

Area	Mastery Learning	vs	Conventional
1. Instructional Basis	-- initially group based but focuses on needs of individual students		-- group based instruction only, does not meet individual student needs
2. Target for Learning	-- all students achieve regardless of previous achievement		-- only high achieving students are successful
3. Testing Procedures	-- formative tests used for diagnostic purposes to improve instruction -- tests are criterion-referenced		-- testing is norm-referenced
4. Grading Criteria	--objective evaluation criteria		-- subjective and based on a curve
5. Instructional Progression	-- corrective instruction consistent and immediate feedback, opportunity to re-teach and additional time to re-do task		-- no opportunity for corrective instruction --no specific feedback on progress

Antecedent Variables

Prior Achievement

Grades and subjective evaluations from previous teachers and scores from standardized instruments were used to determine prior achievement. All students that participated in this study had been placed at the fourth grade level based on these criteria. In addition most bilingual students had been screened by the district's transition criteria for transition into an English-only curriculum.

Writing Ability

The writing ability of the subjects for this study was minimal sentence level writing skills. Generally, prior to the fourth grade level, the students are taught to write basic sentences in English through Language Arts units; Spanish Language background students are expected to produce (at least at a basic sentence level) writing in the English language. Thus the fourth-grade teachers informally evaluate the writing skill level at the beginning of the academic year for ability placement.

Context

Grade Level

The site for Phase I included two fourth grade classrooms selected because of their composition; they were composed of already assembled groups consisting of monolingual English speakers, monolingual Spanish speakers, and a few bilingual students.

The grade level used for this study was the fourth. Fourth grade was selected because at this level there seems to be a shift from task involvement (effort) to comparison (ability) (see Nicholls, 1978). In addition, by the time fourth grade is reached emphasis is no longer placed on students learning the basic procedures for reading and writing but rather on reading and writing for the acquisition of knowledge. Students were evaluated relative to the norm and rarely for individual attainment of skill. Generally, there are a number of presumptions on the part of the teacher: they assume that students have the prerequisite skills to learn more advanced skills.

Subjects

Subjects for both phases of this study were fourth-grade students (n=84) selected from two schools in a community with an integrated population. In Phase I (n=56) there were two teachers assigned to teach fourth-grade language arts. At the beginning of the academic year they meet to determine, on the basis of ability, where students should be placed. Even though all the students were fourth graders, they were divided into two English language ability sections: "high" and "low" achievers. Within these ability groups, there were considerably diverse levels of linguistic as well as academic ability.

In Phase I, students from both ability section groups served as subjects for this study. The teachers use the CTBS (California Test of Basic Skills), grades, and comments from previous teachers to group students into ability sections. In addition, they administer a writing task. The topic for the writing task is taken from the adopted state textbook which stresses elements of paragraph structure and instruction (i.e., indentation, punctuation, capitalization, usage, etc.) for fourth-grade classrooms. The evaluation of the writing samples is done on an informal basis by all the teachers using a holistic approach, samples are ranked as high, medium, or low levels. Finally, Spanish speaking students are introduced to an English-only curriculum at the fourth-grade level. The transition criteria developed by the school district was used at the third-grade level to transition all limited English students. This transition into English-only instruction was done before the students were placed at the fourth-grade and before the initiation of Phase I of this study.

The directions of the Pre and Posttests asked the student to write a paragraph about what a particular holiday meant to them and their friends and family. Moreover, the instructions for Phase I tried to elicit a response by asking them to describe activities or experiences dealing

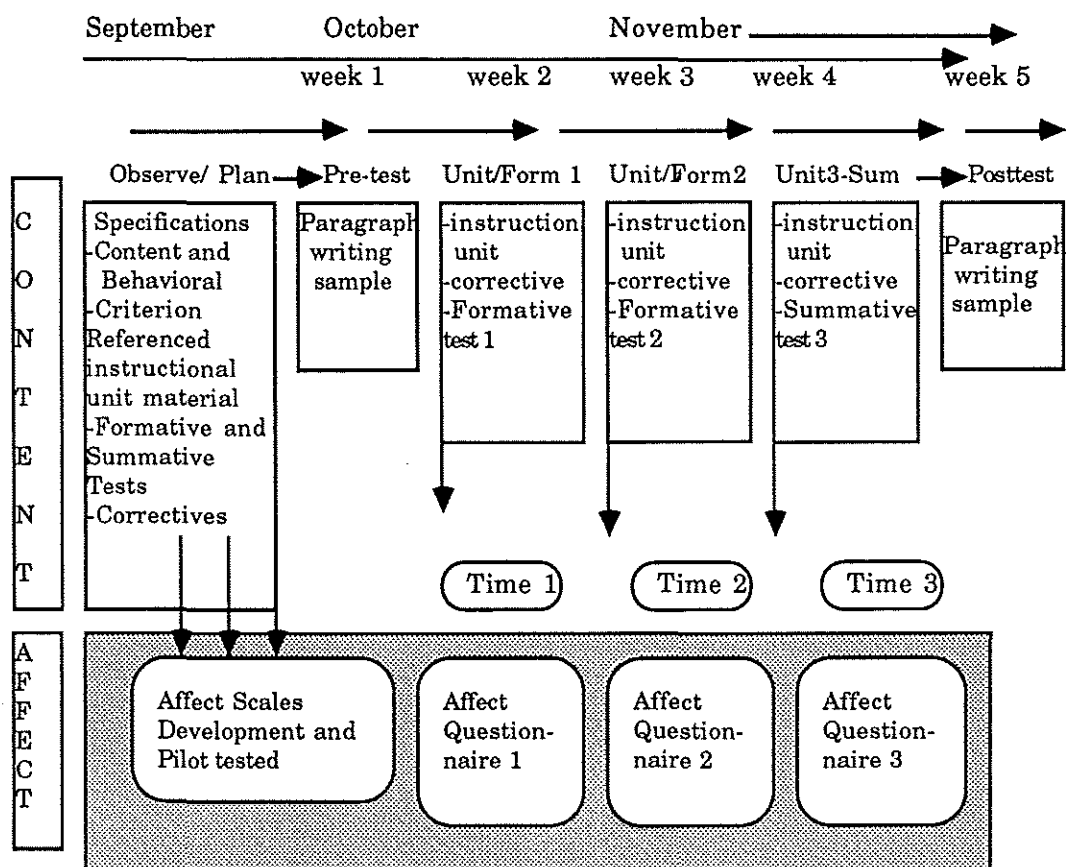
with the holiday in question (either Halloween or Thanksgiving). For the benefit of the students who may have only recently been exposed to these holidays, the directions for the posttest holiday theme (Thanksgiving) asked them to write about what they have learned at school about the holiday's meaning.

Procedures for Data Collection

Figure 3.5 details the schedule of the implementation of the study and the data gathering for Phase I. As illustrated in this schedule, the data collection was conducted by using one grade level and one content area. In Phase I students were taught the same lesson of paragraph structure under the two different instructional conditions. The implementation of the study took approximately five consecutive weeks of instruction.

The actual initiation of the study begun in late August. That was when the researcher first met with the treatment teacher in her classroom to informally discuss and efforts were made to collaborate for proper implementation of the study. At this time, the teacher had already changed her opinion about Mastery Learning. She gave the researcher a copy of the textbook she intended to use and several copies of supplemental material. She did not want to take the time to do anything as a team. Even while the researcher was in her classroom, she continued to put up bulletin boards, file folders, arrange furniture etc. It seem as if she was avoiding the issue of planning for the implementation of the Mastery Learning approach in her Language Arts class period. However, she did answer the researcher's questions concerning student placement and prior achievement and suggested an approach for including the participation of the control classroom teacher (at this time the control classroom teacher had not yet agreed to participate). She also made available the test scores of the students that were enrolled in her class.

Figure 3.5 Schedule of Study Procedures Phase I



It was agreed in early September that the implementation of the study would begin in October since the month of September was filled with chaotic activities (i.e., student placements, aide assignments and teacher/aide planning). These circumstances allowed the researcher to observe, plan the schedule, select and prepare the instructional materials and also to prepare and pilot test the affect questionnaire.

By mid-October everything was ready, and we administered the Pretest and the first questionnaire. The first day of instruction the teacher looked over the unit by unit plan that delineated the instruction, the materials, and the group format. The teacher conducted whole

group instruction during a thirty minute period on word and sentence comprehension and then everyone worked individually. As the teacher presented the lesson that described actions or ideas that were abstract, she wrote words relating to senses or emotions on the chalkboard. The whole group was asked to write sentences using the words that had been generated on the chalkboard.

Individual feedback was provided by the teacher and the instructional aide. The individual feedback was not given on specific concepts, in writing, or for practicing the same concept again, but rather it was given orally and errors on the exercises were not specifically explained. The feedback was more like coaching. The instruction simply progressed from one level to the next. There was no attempt at providing the Feedback and Corrective Loop.

The same procedures were followed from one week to the next and by the end of the month of October and the second week of the implementation of the study the situation worsened; there had to be constant compromise. For example, it was very clear that Formative Tests and Correctives were not going to be used to diagnose levels of learning in order to modify instruction and meet individual learning needs. Although an alternative Formative Test I was developed (for use after corrective instruction and before progressing to the next level) it was not administered.

It was evident that the teacher considered these activities as practice exercises and was not willing to use the results to provide feedback. She considered the endeavor contrary (in her opinion) to the Madeline Hunter approach to teaching. She stated quite openly that she had heard from other teachers that a Mastery Learning approach to teaching was a lot of paper shuffling. Overall, the teacher did not want to hear anything about the philosophical base of Mastery Learning and set out to prove that she was a good teacher in the way that she chose to teach, essentially through a conventional approach to teaching. Thus, by the first week of November, the researcher acquiesced to observing the

teacher deliver the lessons, collecting, and evaluating the correctives (these were used in the form of learning activities) and Formative Tests. The researcher also delivered and collected the questionnaires that students responded to once a week after the Formative Tests. The final instructional unit was completed by the third week in November. The Posttest was administered the the end of the third week in November shortly before the Thanksgiving holidays.

The commitment of the control classroom teacher (in early September) had set the basis for the implementation of the study. However, the control classroom teacher agreed to participate only if it did not require any extra time for planning, evaluating, or conference with the researcher. Later, after the study commenced, this teacher agreed to have the researcher observe the class, but only three times.

In the control classroom, a test or writing sample was collected each week and evaluated by a method that was decided on by the participating teacher and the researcher. Achievement was measured therefore, by the scores on Pre and Posttests and by the number of correct responses on Formative and one Summative Test. In the Mastery Group Phase I, correct responses were compiled from each Formative Test and one Summative Test. Alternate forms of the Formative Tests administered to students who did not initially meet the set criterion level for a given task were not used in this phase of the study. Only the final scores on Formative Tests were used after correctives. In the Conventional Group, the scores on tests that were presented as quizzes were used to gauge achievement.

Since overall achievement for the three instructional units was assessed by a Summative Test in the Mastery Learning classroom, the same form of the test was used in the control classroom as a regular final test. Hence, both groups in Phase I were evaluated by similar criteria. All tests were corrected and scores were tallied immediately after they were administered.

Informal observation was conducted a minimum of four times per week in the treatment classroom. From the vantage point of the researcher, since all of these activities were taking place on an ongoing basis there was hope that once everything was in place the attitude of the treatment teacher would change. She reviewed everything that the researcher developed to see if it was appropriate for the group of students that comprised the class. As the weeks passed, however, her attitude did not seem to change, and, in fact, she seemed more resentful with every passing day. Under these circumstances it seemed as if the researcher was working in a vacuum; there seemed to be only superficial rapport and the teacher went through the motions only because she had agreed to participate. During this entire planning phase, neither of the teachers had any time to meet with the researcher. So the recess (while both teachers supervised playground activities) period was the only time when the researcher met with both teachers together and answered questions concerning the project. Often times the researcher talked to the treatment classroom teacher during her lunch break but this was awkward since she stated on several occasions that she preferred not to use her lunch break to do anything for the project. Of special significance to the research project was the fact that the teacher made it very clear that the researcher was not to use school supplies nor xerox equipment for purposes of the study. Hence everything had to be prepared away from the school site and this became a burden as the study progressed. This was a problem when one considers the amount of xeroxing that had to be done on a daily and weekly basis, aside from the fact that at times it was necessary to make last minute changes on a form or make additional copies of the material. There were a few occasions when some of the materials were xeroxed and a "broken-down" typewriter which was located in the teachers lounge was used but this support was minimal and created an awkward situation for the researcher. Teachers at this school site were

constantly concerned with paper shortage and amount of xeroxing, thus they were all frugal with their use of supplies.

To reiterate, by the week of October 22 (the second week after the implementation of the study), the treatment classroom teacher appeared to be functioning under a great deal of anxiety. The instruction was constantly geared down to provide what the teacher considered "comprehensible input." The teacher as well as the instructional aide constantly commented to the researcher that the students could not be expected to write paragraphs in English since most of them (80%) were Spanish language dominant students and their English language reading scores were lower than most fourth graders. The instructional aide, who occasionally worked with the students (from ongoing observation during the five week period), said to the researcher that she felt that the Spanish dominate students were "language handicapped" since (in her opinion) they could not possibly succeed like the "regular" fourth-grade students.

The fourth-grade was when these Spanish language dominant students were transitioned into an English-only curriculum. The teacher, as well as the instructional aide, at times expressed feeling that if the students reading levels in English were low (barely at grade level according to standardized instruments) they could not learn to write paragraphs in the English language. Thus the instruction focused on word recognition skills and writing basic sentences. In Phase I of the study this was the instructional focus in spite of the fact that the objective of the Mastery Learning instructional units was to teach paragraph structure.

CHAPTER IV

Methodology: Revised Study Phase II

Modifications of the Initial Research Design

In conducting research in educational environments, it is often the case that factors beyond the control of the study change the original proposed design. The major hurdle in this study was the attitude of one classroom teacher toward Mastery Learning which affected its proper implementation in Phase I. Consequently Phase II was implemented.

As we have seen, the negative attitude of the Phase I teacher developed slowly starting in August prior to the onset of the academic school year. When this teacher initially volunteered (during the Spring of the previous academic year), it was with positive intentions. In August however, it became clear that additional time and effort was necessary to plan and develop the basic elements of the Mastery Learning program. The researcher continued to plan and develop the program without the consistent participation of the teacher hoping that her attitude would change by the time the academic instructional sequence commenced. Unfortunately, her attitude did not change and, in fact, by the time October came about there was strong resistance toward the program implementation. Still, she reluctantly agreed to follow the original plan throughout the month of October. At the end of October, the researcher approached the teacher and attempted to ease her out by suggesting that the participation might be curtailed, but she refused and expressed her desire to continue with the project (see time/week discussion of previous Chapter III).

The dissertation chair was consulted and it was decided that a backup site needed to be identified where a presumably less problematic, but conceptually similar implementation of a Mastery Learning approach might be tried. However, since a lot of effort, time, and resources had already been channelled into the original site for the

study we decided to continue there with hopes that something could be salvaged.

A flyer was therefore circulated in the local school district requesting a teacher volunteer to implement a Mastery Learning program in an already assembled bilingual instruction classroom. When the volunteer came forward, Phase II of the study was conceptualized.

The Research Model

Throughout the implementation of Phase II there was a concerted effort made to adhere to the originally proposed theoretical model, hence the same procedure for identifying objectives, developing material and instruments was used in Phase II as in Phase I. In addition, the hypotheses for the overall study were the same in both phases. However, the contextual conditions and variables of Phase II of the study were different. So, even though the procedures and components of Phase I and II remained the same this fact prevents any direct comparison of variables. Figure 4.1 illustrates the theoretical model for Phase II of the study:

Figure 4.1 Schema for Phase II:

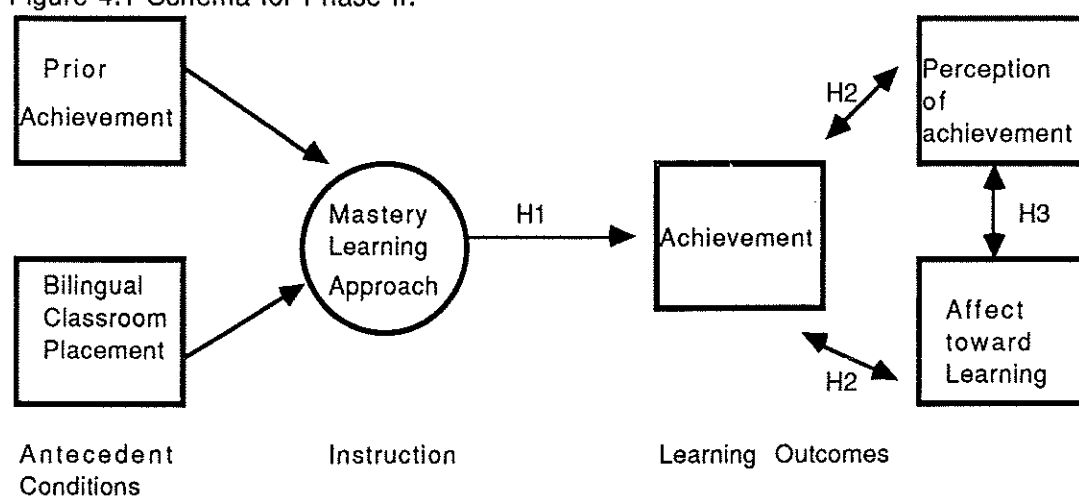


Figure 4.1 Model of Phase II

Research Hypotheses

In Phase II of the study it was hypothesized (H1) that level and variation in student achievement from Pretest to Posttest is a function of the instructional approach. Mastery Learning was the only instructional approach used. This approach continually incorporates procedures (Feedback and Corrective Loop) for enabling students to enter new learning tasks with high cognitive and affective entry behaviors.

To repeat, this phase of the study also had the same three hypotheses as Phase I. As in Phase I, skill in writing paragraphs increases with Mastery Learning and within group variance decreases (H1). It was also hypothesized that Perception of Achievement (H2) and the Attitude and Interest Toward Learning (H3) increases with Mastery Learning. Now, though, each could be tested with one instructional rather than comparatively tested between two approaches.

The Research Design

Unlike Phase I of the study which consisted of a treatment and a control classroom, Phase II of the study only consisted of a Mastery Learning Treatment classroom; there was no comparison group to use as a control. Thus whereas Phase I can be identified by Stanley and Campbell as a Quasi Experimental Pre and Posttest Non-equivalent Control Group Design (Design 10), Phase II is defined as a Pre Experimental One Group Pre and Posttest Design (Design 2).

The following illustration sketches the design of Phase II of the study:

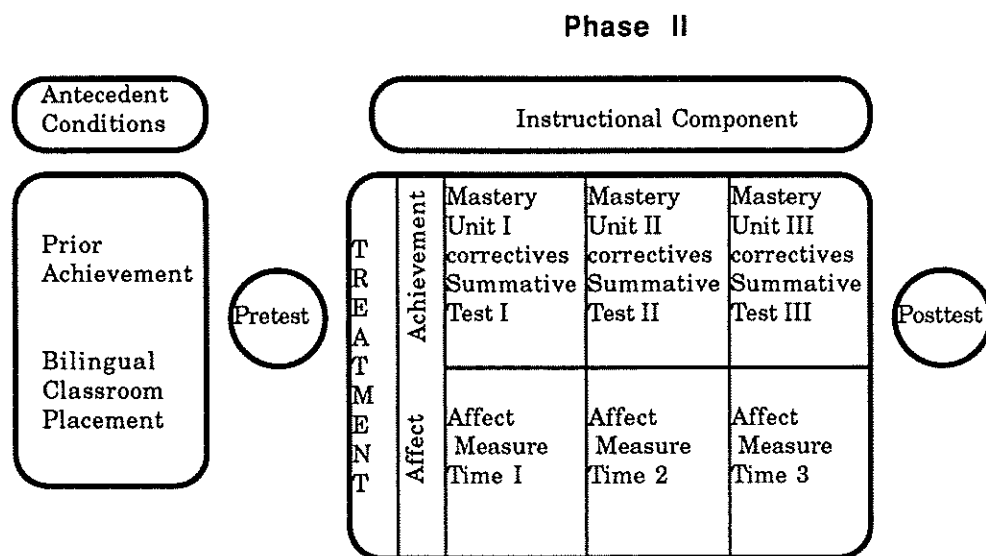


Figure 4.2: One Group Pre and Post Test Design
Stanley and Campbell Design 2 (1963)

Since this design is considered to be a Pre Experimental design it does not account for extraneous variables which may influence the results. The internal validity of such a design is questionable. This design however is an improvement over a one shot case study (Stanley and Campbell Design I) because it utilizes a Pre and a Posttest, thus subject learning gains are measured rather than just how well the subjects did at the end. Unfortunately, without a control group to use as a comparison, no justifiable claims can be made about the effect of the instructional treatment. It was felt however, that whatever could be seen with respect to the implementation of a Mastery Learning approach to teaching writing to bilinguals would complement the findings of Phase I of the study.

Developing the Dependent Variable

Achievement, the primary dependent variable, was measured in the same two ways as in Phase I of the study. Motivation was gauged by the same instrument as in Phase I and was also administered three times.

Gauging the First Dimension of the Dependent Variable: Academic Achievement CSE Subscales and Trait Scales

This procedure was the same as Phase I, except that the dependent variable Academic Achievement was, assessed by using a Mastery Learning approach condition. The same assessment procedures were used, however.

The Measurement Scales CSE and Primary/Secondary

The writing assignment used to generate the Pre and Post writing Samples for Phase II, as in Phase I, was also determined by the same factors, except for the subject matter. These were the concreteness of the topic, familiarity of the themes and ideas by the entire group, maturity level, knowledge of concept, and socio-economic level. In this phase however, the subject matter was different. The teacher felt that she wanted to use a visualization writing task in order to elicit the paragraph writing sample. By observing the approach that she used to manage the class and her attitude toward the students, it was apparent that she had used visualization for storytelling and for generating ideas for writing assignments. Free and creative writing was stressed, and even though students were given guidance and specific directions for doing assignments, there was always time for additional creative writing. In their free time, when students finished assignment rapidly, students could draw pictures about what they had written on the backside of any paper they had used to write. If they chose to write more,

they could write about pictures exhibited on the bulletin boards (which were consistently changed) or the pictures in many of the supplemental reading books that the teacher had displayed in various sections of the classroom. Moreover, students could choose to write in either English or Spanish.

The writing assignment used to elicit the Pretest Writing Sample was a visualization task with three specific parts. The students were asked to write a story with three paragraphs; part one was to describe the meadow then to describe a special animal and finally to tell what the animal told them about itself. They were also told that each paragraph had to have at least four sentences. The same assignment was used to generate writing samples for the Posttest. Creative writing was stressed, but at the same time the notion of topic sentence, indentation, and sequencing were introduced in the unit lesson sequence and continually reinforced by the feedback and correctives provided.

Gauging the Second Dimension of the Dependent Variable: Academic Achievement

The assessment procedures were, again, the same as in Phase I except that there were more Formative Tests used, and only the Summative Tests were used as Final Achievement Measures.

Gauging the Third Dimension of the Dependent Variable: Motivation

The procedure for gauging the third dimension was exactly the same as in Phase I. The Perception of Achievement, Attitude, and Interest questionnaire was administered three times during a five week period.

Developing the Instrumentation Phase II Components

CSE Scale for Evaluating Writing

The evaluation of the writing sample for Phase II was exactly the same as in Phase I. All writing samples for both phases were in fact evaluated at the same time.

Primary/Secondary Trait for Evaluating Writing

The evaluation procedure established for use in Phase I was the same procedure used in Phase II.

Formative and Summative measures

Unlike the Formative Tests in Phase I, the Formative Tests in Phase II actually served to identify the areas where students needed additional practice and correctives. In addition, Correctives served as enrichment activities for students who did not need the practice. The Formative and Summative Tests were part of the instructional sequence and coincided with the Table of Specifications (see Appendix A). In Phase II, therefore, the Summative Tests were scored for Achievement results.

Summative Test I Phase II

Summative Test I required that the students identify the topic sentence in a sample paragraph. They were given three sample paragraphs and asked to rewrite them and leave out the sentence that did not belong. Sentences that did not belong to the content of the paragraph were used as distractors. This test had a total of three correct responses. The following is a sample item from Summative Test I:

Sample Item Summative Test I English version:

All sentences in a paragraph should tell something about the main idea of a paragraph. Rewrite each paragraph below, leaving out the sentence that does not belong.

Lighthouses help ships pass dangerous places. At first people used fires built on top of high hills. Big boats can carry many people. Today lighthouses use bright light to do the same job.

Spanish version:

Todas las frases de un parrafo deben contar sobre una idea principal del parrafo. Re-escribe cada parrafo que sigue, pero no uses la frase que no pertenece.

Los faros ayudan a las naves a pasar por lugares peligrosos. Al principio la gente usaba lumbres construidas sobre cerros altos. Naves grandes pueden cargar a mucha gente. Ahora los faros usan luces alumbradas para hacer el mismo trabajo.

=====

The content of Summative Test I (Sum 1.1) reflect the lesson content and the corrective activities that students did which were designed according to the diagnostic results of the Formative Tests .

Summative Test II Phase II

In Summative Test II students were required to indent. They were given a group of sentences and were asked to divide them into three paragraphs and then indent the first sentence of each paragraph. This section coincides with section 1.4 of the Table of Specifications (see Appendix A). Again there were a total of three correct responses for this Test.

The following is a Sample Item from this test:

English version of Summative Test II:

The first sentence of a paragraph is indented. We write it in from the left margin to show where a new main idea begins.

Write the following group of sentences.

Divide them into three paragraphs.

Be sure to indent the first sentence of each paragraph.

Glass is made by heating a special kind of sand. Plain glass is always green. But things are added to take away the color or to make other colors. There are many ways to make something out of glass. Big pieces of glass can be cut.....

Spanish version:

La primera frase de un parrafo es endentada. La escribimos algunos espacios del margen de la izquierda para enseñar en donde comienza una nueva idea.

Escribe el grupo de frases que sigue.

Dividelas en tres parafos.

Ten seguridad de endentar la primera frase de cade parrafo.

Para hacer el vidrio se calienta una arena especial. El vidrio natural siempre es verde. Pero se le agregan cosas para quitarle el color o para hacer otros colores. Hay muchos modos de hacer algo de vidrio. Pedazos grandes de vidrio se pueden cortar

=====

The content of Summative Test II (Sum 1.2) likewise reflects the progressive building of knowledge assumed to be gained from the diagnostic process used for Unit I and assessed by Summative Test 1.1.

Summative Test III Phase II

Summative Test III assessed all the skills for all parts of the four week instructional sequence. It required the student to write three separate paragraphs, sequence the sentences, and identify the topic sentence by underlining. There were a total of three correct responses coinciding with the cluster of objectives in Section 1.5 of the Table of Specifications (see Appendix A). The following are sample items of Summative Test III:

English version of Summative Test III:

Write this paragraph, putting the sentences in the right order.

Leave out the sentence that does not belong.

Put a line under the topic sentence.

It's easy to make different colors with paint. Then you can
make all the other colors by.....

Write the following group of sentences.

Divide them into two paragraphs.

Be sure to indent the first sentence of each paragraph.

Gold is a soft, yellow metal. It is twenty times heavier
than water. It is easy to shape and.....

Spanish version of Summative Test III

Escribe este parrafo, pon las frases en el orden correcto.

Suprime la frase que no pertenece.

Subraya la frase del tema.

Hacer diferentes colores con pintura es facil.
Luego puedes hacer todos los otros.....

Escribe el grupo de frases que siguen.

Dividelas en dos parafos.

Ten seguridad de endentar la primer frase de cada parrafo.

El oro es un metal blando y amarillo. Es veinte veces mas
pesado que agua. Es facil formarlo siempre.....

=====

The content of the third and final Summative Test (Sum 1.3) assesses a synthesis of skills that were presumably acquired through a continual feedback and corrective process throughout four instructional units used in this bilingual Mastery Learning teaching approach.

Questionnaire to Test Motivation

The questionnaire used to test motivation was exactly the same form as in Phase I and the same procedure was followed.

Developing The Independent Variable

The independent variable in Phase II of the study was the instructional approach. In this particular Phase only a Mastery Learning Approach in a bilingual instructional class was used.

The Mastery Learning Approach

The Mastery Learning Approach used in Phase II was the same as in Phase I but the objective was not to compare two learning methods rather, the objective was to test the effectiveness of a one group Mastery Learning Approach for teaching basic writing to bilinguals in a single condition situation.

Antecedent Variables

Prior Achievement

As in Phase I, grades and subjective evaluations from previous teachers and scores on standardized instruments were used to determine prior achievement. The basic difference from Phase I was that the prior bilingual achievement of the students in Phase II was used as a criteria for the opportunity to participate in a fourth-grade bilingual classroom.

Placement in the bilingual classroom was not mandatory but rather voluntary, hence all monolingual English students and bilingual students (various levels of Spanish/English bilinguality) volunteered. The parents of these students could choose to enroll their son or daughter in an all-English instruction fourth grade classroom but chose the bilingual setting. Bilingual standardized tests were used to assess linguistic proficiency in both languages. Most of the monolingual English speakers were not fluent in Spanish, but all material and instruction at all levels were offered in English as well as in Spanish.

Thus bilinguality was seen as an enrichment while at the same time providing for basic skill acquisition.

Context

Grade Level

Phase II of the study included a fourth grade classroom at a different school than Phase I. Instruction was bilingual (Spanish/English), and students were given an opportunity to choose the language group for instructional purposes. Since the level selected for Phase II was also the fourth grade, the same rationale for learning that was used in Phase I held true for Phase II. Students seem to shift from task involvement to comparison of ability (see section for Phase I components). Phase II differed from Phase I however in that this was a bilingual Spanish/English instruction classroom. Various levels of English language proficiency did not seem to make a difference in the way the groups were composed; there were various levels of linguistic ability in both languages. All students, however, were included in the overall instruction since it was presented in both languages and the lessons seem to flow smoothly for all involved.

In Phase II, the teacher selected the materials to be used and specified the performance objectives for four weeks of instruction. The researcher then developed the Table of Specifications. This was done by using the daily instructional objectives that the teacher wrote. The researcher selected the Formative and, in this case, Summative Tests also from the material selected by the teacher. There was an attempt to correlate the performance objectives in Phase II to the performance objectives in Phase I. Even though the language of the objective for both phases is different (since the objectives for Phase II reached a higher level of abstraction), the basic goals for the four weeks of instruction in both situations were the same; the development of a paragraph. All of

the materials for instruction, including the Formative and Summative Tests in Phase II were translated into Spanish by the researcher.

Also unique to this situation was the involvement of the researcher as an instructional aid during the implementation of the study. This was one of the prerequisites agreed on by the researcher and the teacher at the onset of the study. As is common in most bilingual instruction classroom situations, in order to insure that all of the material is covered in both languages, instructional aides are used to assist the teacher. In this particular situation, the instructional aide who normally assisted this bilingual teacher was not available for the time period allocated for writing instruction. It was crucial, that the researcher participate in the proper implementation (which included instruction) of all the facets of the Mastery Learning Writing Sequence. The researcher, therefore, was not only helping the teacher to plan, identify appropriate material, translate, type, and prepare all material, she was also assisting in the instruction and evaluation. This constituted an instructional research team in a site that can appropriately be identified as a bilingual instruction classroom.

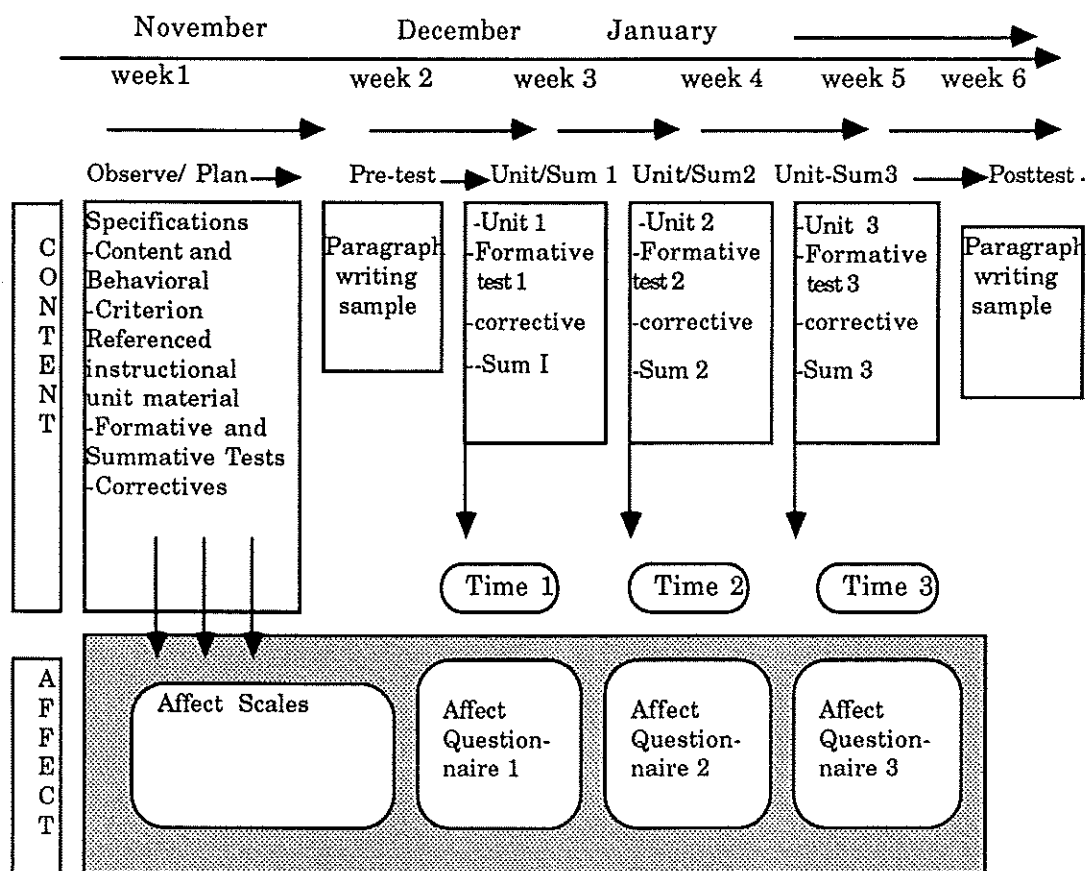
Subjects

Phase II subjects (N=26) were students assigned to grade four primarily on the basis of test scores and grade level scores on standard instruments. In addition, the results of standardized scores on Spanish language Tests were used to develop bilingual Spanish/English curriculum (this was done on an ongoing basis with the primary responsibility given to the classroom teacher). The subjects used for Phase II were not channelled into ability sections. They were intermittently grouped for instructional purposes in order to provide instruction in one language. Bilingual students (the ones that seemed comfortable in either language) were given the choice of what language to use for learning. The language of instruction did not seem to be a

problem since the same instructional materials were used with both groups. The only time the class was divided into groups was after the main lesson presentation. Then the instruction was conducted in groups until the entire class reached the desired level of performance. During this time the students were provided with various levels of materials to reach the same objectives.

Procedures for Data Collection

Figure 4.3 Schedule of Study Procedures Phase II



As illustrated by Figure 4.3 schemata of procedures, the data collection for Phase II was similar to that of Phase I. One grade level and one content area was used. The students in Phase II were taught paragraph structure under a bilingual Mastery Learning approach condition. The implementation of this phase of the study took approximately six consecutive weeks of instruction during three months (since this phase started at the end of November the sequencing of activities was interrupted by the Christmas holiday vacation).

During the planning phase, observation was conducted in two ways: informally and formally. The informal observation was done during the second to last week of November (the researcher first met with the teacher before the Thanksgiving holidays). The teacher insisted that the researcher come to the classroom daily so that the students could familiarize themselves with a new adult person. In addition, this was the time when the researcher met the principal, secretaries, and instructional aides and found the resources available to use for the study (i.e., xerox machines, typewriters, mailboxes). The use of supplies and office machines was not a problem at this site. Everything was made available and in a very congenial manner.

The last week of the month of November was used to formally observe the classroom proceedings and to develop the Table of Specifications, identify and develop the criterion reference measures (i.e., Correctives, Formative and Summative Tests), and organize the instructional units. Once everything had been organized in English, the ongoing translation begun (this was done at home by the researcher and each lesson unit was checked on a continual basis by the teacher). By the end of the final week in November, the Pretest was administered and also the first Perception of Achievement Questionnaire.

The teacher officially started the instructional units on the first week of December, but there were so many interruptions because of Christmas (parents visiting and school holiday activities) she felt that we could continue the unit the second week in January. By the end of

the first week in December, however one instructional unit had been used along with correctives, Formative Tests and one Summative Test administered. This same plan was followed throughout the month of January. The final Posttest was administered on January twenty-fifth.

In this Mastery Learning group classroom students were taught as a whole group. After the initial lesson presentation in both languages, students worked in separate language groups, either with the teacher or the instructional aide. The teacher and instructional aide often switched groups, in other words the instructional aide did not only work with the bilingual students or vice versa. Also the bilingual students often switched groups; there were some students who were sufficiently proficient in both languages and often worked in either group. Students were encouraged to work with peers within the same group or with members of the other group (i.e., monolingual English speakers with bilinguals or monolingual Spanish speakers). Students who needed practice were given a variety of techniques in order to practice the same concepts. Similarly, the students that needed enrichment activities were provided opportunities to further express themselves in writing and oftentimes these included reading and writing together. It was often necessary to extend the class time beyond the regular thirty-minute class period for students that needed individual attention. However, this did not ever extend beyond fifteen minutes (this happened to be the lunch break for both students and teacher). When these students needed the extra time, both the instructional aide and the teacher stayed to supervise and work individually with those students who stayed to finish the task. The extra time allocation occurred often during the six week instructional sequence.

By following this routine, ongoing achievement was measured by the scores on Summative Tests which were indicators of a synthesis of skills that had been continually practiced with corrective activities and

diagnosed by Formative Tests . All tests were corrected and scores were tallied immediately after they were administered by the researcher. The teacher simply followed the instructional sequence and reviewed the results.

CHAPTER V

Data Analysis

This study sought to investigate whether learning, interest, and Attitude Toward Learning could be altered positively for students of Spanish language background regardless of placement or prior achievement by using an equity structure approach. The study to repeat, had two phases.

In Phase I, the cognitive and affective learning of Spanish language background students was investigated under two learning conditions: conventional group instruction and a particular equity structure, Mastery Learning. It was hypothesized that the student achievement would be a function of the instructional approach used and that student learning would be higher under Mastery Learning than under conventional instruction. Moreover, it was hypothesized that Attitude and Interest towards learning develop as a function of the achievement students attain and of their perception of the adequacy of their achievement. Since Mastery Learning was hypothesized to generate greater achievement than conventional instruction, the Attitude and Interest in the Mastery Learning Groups was expected to be higher than that found for the conventional group.

In Phase II, the cognitive and affect of Spanish language students was further investigated in only one instructional condition, a bilingual, Spanish/English, equity structure condition. Again this particular equity structure was Mastery Learning. In this phase, it was hypothesized that level and variation in student achievement from Pretest to Posttest would be a function of the instructional approach. Specifically, it was expected that skill in writing paragraphs would increase with Mastery Learning and within group variance in this writing skill would decrease. As in Phase I, the increase in paragraph

writing skills should have influenced Perception of Achievement and the Attitude and Interest toward learning.

Scoring, Reliability and Validity Checks

In this section the procedures for establishing reliability and validity of the instruments used for this study will be described. First it is necessary to describe the initial considerations taken into account so that all activities for the analysis of the data were dealt with in a smooth and thorough manner. The scoring procedures as well as the coding of the raw data are explained in separate sections since these activities set the basis for conducting further analysis.

Pre-Post Test Writing Sample Scoring: Procedure

The entire pre/post evaluation process lasted two months. All of the writing samples for the three groups included in both phases of the study were rated by the same procedures and the same evaluators. First, the raters were recruited, interviewed, and trained for two weeks. During this time, effort was made to establish conformity between the raters (see pg. 51-53 of Chapter III). The conformity of the rating process was accomplished by allowing two bilingual raters to score fourteen writing samples of various levels and in both languages by using the selected modified CSE Subscales. The scores were then scanned and if there were any discrepant scores (one score above or below), the samples were re-scored by the third rater and the researcher. The discrepant scores were then discussed until a consensus was reached. Also, the language of the scales was further clarified by giving specific examples that constituted various levels from the samples. Writing samples used for these pilot test sessions were selected samples that study subjects had written but that were excluded

from the final samples taken for statistical analysis. For instance, if a particular subject had the Pretest but not the Posttest or vice versa, these samples served the purpose of the pilot sessions well since only complete (Pre and Post) tests were used for the final analysis.

All raters used the selected CSE Expository Writing Subscales to score the writing samples over three consecutive weeks for approximately fifteen hours each week. Because, as previously mentioned, each writing sample was scored by two different readers, the raters checked their ratings and responses to master score sheets on which score rationales were described. Every three to four hours, the scores for each group were checked and monitored by the researcher.

The two readers also rated the same writing samples with the Primary and Secondary trait rubrics. The training for the trait scales lasted three hours and the actual rating lasted five hours. When scores differed by more than one point, a rating was done by a third reader. If the sample was written in Spanish, it was rated by the researcher; if not, it was rated by a alternate reader who was on-hand as an ongoing consultant.

Establishing Reliability in Scoring

CSE Subscales and Trait Scales

The raters who conducted the Pre and Posttest evaluations had absolutely no connection to the instruction or the data gathering phase of the study. They were hired specifically to assess the writing samples. In this way objectivity in the assessment could be maintained.

After the raters had been recruited, interviewed, hired, and trained, there was a conscientious effort made to preserve the anonymity of the subjects and to mask the identification of the various groups used in the study. The researcher selected the standards for evaluation, prepared two packets of writing samples and made arrangements to conduct the entire process at CASE (the Center for Academic Skills

Enrichment), UCSB. The packets of writing samples consisted of different sets of Pre and Posttest materials and additional paragraph writing samples that were to be evaluated by a set criteria. During the training, pilot sessions, and the actual rating sessions, randomly arranged writing samples were presented to the raters. They did not get the samples in any logical or sequential order.

The two raters who did the bulk of the scoring had separate copies of the writing samples (a third rater and the researcher were available as alternates). They followed the same schedule and worked in the same room. They were asked not to discuss their ratings with one another until after the debriefing which was conducted and monitored by the researcher at the end of each rating session.

The rating schedule was limited to minimize rater fatigue. Raters assessed writing samples in four-hour time periods and were instructed to get up and stretch, drink water, etc., at any time during their scoring as long as they maintained their focus and train of thought. There were no reported fatigue problems.

Since both raters had different backgrounds yet similar training, there was, at times, disagreement even when using a fixed set of criteria for scoring. For instance, they seemed to disagree on the level of importance of spelling errors; one rater felt that certain errors did not prevent clarity of expression and was more lenient than the other. Another area which caused some confusion was the level of the scoring criteria that constituted "Mastery" vs "Non Mastery." One rater started using decimal points with the score categories of the scales.

The problems regarding scoring disagreements were dealt with by discussing them during the training sessions and at the end of every actual rating session. It was sometimes necessary for the researcher to select samples to show the raters specific examples of the distinct levels of language identified in the subscales.

Table 5.0 presents the interrater reliabilities of all scales used for the assessment of the Pre and Posttest writing samples for both phases

of the study. All interrater reliabilities reported were calculated by using Pearson's Product Moment Correlation with listwise deletion.

Table 5.0

Interrater-Reliabilities for CSE and Trait Scales **

<u>CSE Subscales</u>	<u>Pre</u>	<u>Post</u>
Subscale 1 General Impression	.90	.96
Subscale 2 General Competence	.92	.97
Subscale 4 Paragraph Coherence	.86	.97
Subscale 6 Mechanics	.85	.96
<u>Trait Scales</u>	<u>Pre</u>	<u>Post</u>
Primary Trait	.97	.97
Secondary Trait	.96	.98

** Note: Interrater reliability reported were Pearson Correlation, using listwise deletions.

As can be seen from these correlations all interrater reliabilities were high (.90 or higher), with the exception of two Pretest reliabilities which were at .85 and .86 for the two most analytical scales (subscales 4 and 6). Greater interrater disagreement in scoring was expected here, though, since presumably such scales would be the toughest to score.

Scoring: CSE Subscales

Each student completed the Pretest on the first day and the Posttest on the last day in both phases of the study. These writing samples yielded the Primary Data Set. Each of the CSE Subscales selected for the assessment of study writing samples utilized scores across a six point range. A "1" is the lowest score and a "6" is the highest. The scoring guides (See Appendix C) describe the skill and errors defined by each point of the scoring range. The "mastery" level is described as the difference between a score of "3" and a score of "4." In other words, the score of "4" is the minimum level of that scoring system considered to be competent performance of a given skill. To reiterate on a point in the previous chapter, only four out of six CSE Subscales, CSE Scale IV (refer to Chapter III for a full description of the scale) were

appropriate for this study. Since the CSE scales were originally designed to evaluate expository essays rather than single paragraph writing samples, the language of the subscales was modified to serve the criteria and purpose of the study.

Scoring: Primary Trait

All of the paragraph writing samples for all groups for both phases were also evaluated by a Primary and Secondary scoring rubric. These trait rubrics yielded a holistic score (as explained in Chapter III). A holistic score was obtained for a Primary Trait relating to the main objective of the four week instructional sequence in both phases: the presence or absence of a topic sentence. The Secondary Trait also yielded a holistic score for the presence or absence of inventive expression and elaboration on a topic.

The scoring rubric developed for assessment of the Primary Trait utilizes a five point scale of "1" to "5." A score of "1" was used to identify papers that had one or more sentences that did not relate to the assigned topic and were not coherent in content. A minimum mastery level (score of "4") was given for identifiable paragraphs that had an explicit topic sentence and in which all sentences had a logical sequence. The justification for determining this level as a mastery one was the instructional objectives for both studies: a cohesive paragraph had to have an explicit topic sentence. In all the instructional sequences for both studies, topic sentence was taught to appear as the first sentence of a paragraph as opposed to the last one. This sentence then provided for the interrelationships and subordination of all the other sentences that appeared in the paragraph.

The Secondary Trait scoring rubric also ranged along a five point scale for rating but was by far the most rigid in terms of expectations for mastery level. The score of "1" in these scales was reserved to identify papers that clearly showed incomplete thoughts and made no attempt at

expressing a cohesive idea. A minimum level of mastery is a level "4" which shows the writer's use of experience or imagination to express concrete thoughts. A level "5" allows for the presence of inventive expression through abstraction and elaboration to be considered for the purpose of including these samples in the evaluation.

Scoring: Formative and Summative Tests Phase I

As mentioned in Chapter Three, the Formative and Summative Tests are criterion-referenced; thus, they vary in scope, length, and focus. Correct responses for each section of every test were given one point (see test sample item descriptions Chapter Three). Wrong responses were not counted. Thus, for example, if in Formative Test I, Part A there were twelve correct responses and the student gave six, a score of "6" for Part A was recorded.

Mastery level was determined separately for each test according to the total number of items. Formative Test I had a total of nineteen correct responses and 85% of these correct responses was required for "mastery," so students had to receive a score of "17" or higher on this test in order to "master" the instructional unit for which this test corresponds. Formative test II had a total of "18" correct responses; students had to receive a score of "16" or higher to demonstrate mastery of the instructional unit for which this test corresponds. Item scores were tallied immediately after the the tests were administered and recorded on graph paper with the coded subject numbers.

The Formative Tests used for both groups in Phase I gradually increased in complexity. So, a single, cumulative Summative Test that conceptually integrated the skills presumably gauged by the two Formative tests, was the last test administered before the Posttest. This test consisted of only four correct responses (see Chapter III for sample item description) and was scored along a five point scale; 4, 3, 2, 1, 0. As the case of the Formative test, the final scores on the Summative test

were tallied immediately after they were administered and recorded on graph paper. The maximum score of "4" was considered Mastery and was reserved for papers that sequenced the sentences in a paragraph form (i.e., indented and capitalized the first sentence, did not write out each sentence in a disjointed, manner but actually joined each sentence to give the semblance of a cohesive paragraph).

Scoring: Formative and Summative Tests Phase II

As explained in Chapter III, in Phase I the different versions of the Formative Tests were not used for diagnostic progress purposes. So there was no recourse but to use the available scores from the two Formative Tests and one Summative to gauge periodic achievement. In Phase II, however, there were multiple versions of each Formative Test that were consistently administered and used for correctives and feedback per instructional unit. These served an ongoing diagnostic purpose. So, only the scores for the Summative Tests were used to gauge final achievement for each instructional unit.

These Summative Tests were cumulative in nature since they were developed based on the results of the previously administered Formative Tests and corrective activities for each of the three instructional units. Since these tests were cumulative, they consisted of fewer items than the Formative Tests; there were a total of three items for each test. All three tests were scored in the same manner. Scores were tallied immediately after each test was administered and recorded on graph paper. Finally, the criterion used for judging mastery of skill or knowledge of concepts in each instructional unit, as gauged by each Summative Test item, was absolute; students either did or did not write a cohesive paragraph, indent, or delete the inappropriate sentence, underline the topic sentence, or whatever was required to respond correctly. Each Summative item (cumulative knowledge item) was

either given a score of "1" or "0." A Mastery score for each test was a score of "3".

Scoring: Perception of Achievement,
Attitude and Interest Questionnaire

All three sections of the questionnaire were scored by using the same scale. Students responded to questions that were weighted according to an ordinal scale. The responses were either "yes," "don't know," or "no" to questions for each part of the questionnaire with the exception of items 12 and 13 in Part A which required the respondent to circle the best response. The total number of questions in the questionnaire for all three sections therefore was 26.

Part A of the questionnaire gauged Perception of Achievement. This section was comprised of 13 items that dealt with how the student saw him or herself with regard to class participation, classroom teacher, grades received, and ability in writing. The responses were weighted on a three point scale. The format for 11 items out of 13 was varied so that in some cases a "yes" answer received a weight of 3 and in others a "no" response was weighted as a 3. For instance, the Part A question "I am discouraged with my writing ability" would be weighted a 3 if answered by circling a "no" response. The "don't know" response was weighted a 2 since it was the neutral response. Two items in Part A required the student to circle an appropriate answer for three choices regarding grades and the general level of their academic work. A response of "excellent" or "best grades" for instance, were given a weight of "3"; at the other end of the spectrum, "poor" grades or "poor" work were weighted as a "1" response.

Part B of the questionnaire assessed Attitude Toward Learning to write by using the same format as the first 11 items in Part A. This section consisted of a total of seven items that dealt with Attitude Toward Learning to write. For example, "Writing paragraphs is more like

playing a game than school work" required a "no", "don't know", or "yes" response. The responses were weighted the same as in Part A.

Part C followed the same format as Part B; the items were weighted in the same manner, but their content consisted of Interest towards writing. Part C consisted of six items that dealt specifically with student interest toward different aspects of learning to write. For example, one question stated "I think learning to write sentences and paragraphs is a waste of time." The student could respond "yes", "no", or "don't know."

All responses were tallied at the end of each questionnaire administration (questionnaires were administered three times in each instructional condition) according to the number of "yes", "no", or "don't know" responses for each group. Additionally, each separate response for each item was recorded on graph paper directly from the original questionnaires

Coding The Data

Procedures for Data Coding

Data for the CSE Subscales, Trait Scales, Formative and Summative Tests, and Perception of Achievement Questionnaires were coded and entered by the researcher into data templates on a Compupro multitiser computer system by using Dataflex database software. Three separate files were developed and were then written to IBM P C format disks. All analyses were done by using Systat data analysis software (Version 3.0).

The data yielded from each instrument used in this study was coded separately. Each dependent variable was measured by a separate instrument. Thus for the Pre and Posttest data, there were six dependent variables and six different measures to assess these primary dependent variables. These scores constituted the preliminary data set for both phases of this study. The remainder of the data (Formative,

Summative and Questionnaire data) were gathered during the period of time between the Pre and Posttest (please see procedures for both phases, Chapter III and Chapter IV). A preliminary coding scheme was used on site as soon as the data was gathered. Everything, including the Pre and Posttest samples, was coded according to class (Class 1 control, Class 2 Treatment in Phase I and Class 3 Treatment II in Phase II) and Subject ID number. The subject ID number started with "1" thru "54" for Phase I and continued with "55" thru "84" for Phase II.

Coding the CSE and The Trait Scales

The primary achievement data were those data obtained from the Pre and post measures; the four CSE Subscales, one Primary Trait scale, and one Secondary Trait scale. Two separate scoring matrices, one per rater, were developed by the researcher to record the scores that were obtained from the four CSE Subscales. Each matrix had five columns; one column for a subject ID number and one column for each of the four subscales. Thus, when each individual rater scored a writing sample, the score was recorded in the appropriate column.

The same procedures were followed with the "holistic" measures, the Primary and Secondary Trait scales except that the matrices used to record these scores only had two columns; a subject ID number column and one score column for a single "holistic" score for each writing sample.

To reiterate, all of the writing samples were coded by an ID number when the data were gathered. All the Pre and Posttest samples were coded for both phases of the study from subject number 1 to subject number 84. When the scoring was completed the data were coded according to Subject ID number, Classroom number (this was either control, treatment 1 or treatment 11), CSE subscale or Trait scale used, and rater (either R1 or R2) before they were entered into the computer

data editor program by the researcher. For instance, PRECSE1R1 from the data coding sheet means Pretest CSE Subscale 1 Rater 1 and POCSE 2 R2 means Posttest CSE Subscale 2 Rater 2, etc.

Coding the Formative and Summative Tests

There were two Formative Tests and one Summative used in Phase I. The items for the Formative Tests were coded according to each test section and to the instructional unit to which each item corresponded. For instance, Formative Test I consisted of two parts and corresponded to the first instructional unit; therefore, each item was coded according to these features. For example, FT1U1A1 translates into Formative Test I Unit 1 Part A Item 1 and FT1U1B2 is Formative Test 1, Unit 1, Part B, Item 2. In this manner every item for both Formative Tests in Phase I of the study was coded on grid paper and then entered into the computer. Scores for both the control and treatment classrooms in Phase I were recorded on grid paper according to a Subject ID number and whether they received a one or a zero for each item on each of the Formative Tests. The subject ID numbers ranged from 1 to 54 for Phase I of the study. The one Summative Test for Phase I consisted of four correct responses and was coded and recorded in the same way as the Formative Tests for both classes. Thus, SUM3U3A1 translates into Summative Test 3, Unit 3, Item 1. The responses for each item for both groups were recorded on grid paper, coded, and entered into the computer by the researcher.

The Summative Tests used to gauge periodic achievement in Phase II were coded in a similar way. There were three Summative Tests, three instructional units, and three items for each test. The scores for each response were recorded on grid paper according to a Subject ID number; these ranged from 55 to 84. Similarly, each item for each test and each instructional unit was coded as SUM1U1A1 or

SUM3U3A3 and so on. Each score for each item was then recorded on grid paper and entered into the computer.

Coding The Perception of Achievement , Attitude and Interest Questionnaire

Ancillary data from the administration of a three part questionnaire to gauge motivation towards learning was operationalized as A) Perception of Achievement, B) Attitude Toward Learning and C) Interest in Learning to Write. These variables were coded in the same manner for both Phases of the study. Every item on each questionnaire for each administration was coded before entering the scores into the computer data files. Since the questionnaire consisted of three parts each item was coded according to each corresponding section of the questionnaire and to each time of administration (see previous section on scoring, each questionnaire was administered three times for each group). For instance, Att1A1 translates into Attitude 1, Part A, Item 1, and Att3B7 translates into Attitude 3, Part B, Item 7. Thus, each questionnaire was coded according to the time of administration (1, 2, or 3) and also to each section and each item. Therefore, for each group, a subject ID number and responses were coded on separate grid matrices before the data were entered into the computer templates.

Summary Sample of Variable Codes

Table 5.1 illustrates samples of the coding schemes for each of the instruments used in the study. Included in the CSE Subscales and Trait Scale sample schemes, are the codes which identify each rater.

Table 5.1

Samples of the Instruments'

Coding Schemes

CSE AND TRAIT SCALES SAMPLE CODING SCHEMES

PRCSE1R1=	PRETEST CSE SUBSCALE 1 RATER 1
PRCSE1R2=	PRETEST CSE SUBSCALE 1 RATER 2
PRCSE2R1=	PRETEST CSE SUBSCALE 2 RATER 1
PRCSE2R2=	PRETEST CSE SUBSCALE 2 RATER 2
POCSE4R1=	POSTTEST CSE SUBSCALE 4 RATER 1
POCSE4R2=	POSTTEST CSE SUBSCALE 4 RATER 2
POCSE6R1=	POSTTEST CSE SUBSCALE 6 RATER 1
POCSE6R2=	POSTTEST CSE SUBSCALE 6 RATER 2
PRPTSR1=	PRETEST PRIMARY TRAIT SCALE RATER 1
POPTSR2=	POSTTEST PRIMARY TRAIT SCALE RATER 2
PRSTSR1=	PRETEST SECONDARY TRAIT SCALE RATER 1
POSTSR2=	POSTTEST SECONDARY TRAIT SCALE RATER 2

FORMATIVE/SUMMATIVE SAMPLE CODING SCHEMES

FTIU1A1=	FORMATIVE TEST 1 UNIT 1 PART A ITEM 1
FTIU1B1=	FORMATIVE TEST 1 UNIT 1 PART B ITEM 1
FT2U2A1=	FORMATIVE TEST 2 UNIT 2 PART A ITEM 1
FT2U2B1=	FORMATIVE TEST 2 UNIT 2 PART B ITEM 1
FT2U2C1=	FORMATIVE TEST 2 UNIT 2 PART C ITEM 1
SUMM3U3=	SUMMATIVE TEST 3 UNIT 3 PART A ITEM 1

ATTITUDE QUESTIONNAIRE SAMPLE CODING SCHEME

ATT1A1=	ATTITUDE QUESTIONNAIRE 1 PART A ITEM 1
ATT1B1=	ATTITUDE QUESTIONNAIRE 1 PART B ITEM 1
ATT1C1=	ATTITUDE QUESTIONNAIRE 1 PART C ITEM 1
ATT2A2=	ATTITUDE QUESTIONNAIRE 2 PART A ITEM 2
ATT3C3=	ATTITUDE QUESTIONNAIRE 3 PART C ITEM 3

Establishing Reliabilities

The values of the scores for the Pre and Posttest CSE Subscales is tempered in part by the fact that the wording of the scales was modified in order to be used for this study (see Chapter III and Appendix C). The original CSE Expository Scale IV presumes the existence of generalizable features of good writing, regardless of the topic or audience of the given essay task. The basic writing task for this study, however, was not an entire essay but a paragraph. Therefore, the selected subscales were modified as noted to serve this purpose.

The basic analytic and holistic categories of the CSE expository subscales can be used for a variety of expository tasks; the Primary and

Secondary Trait is a task-specific, "holistic" scoring method. The scoring rubrics have been built from a careful analysis of the features of the proposed writing task--the basic units of a paragraph with an explicit topic sentence as the main idea. This second rating method used to score the Pre and Posttest in this study presumes just the opposite of the CSE Subscales, that writing is highly affected by the context and content and other features of the task.

Cronbach's Coefficient Alpha was used as a measure of score reliability for the different instruments used in the study. Iterative procedures were used to estimate the maximum possible reliability. Pearson Product-Moment Correlation Coefficients and related Principal Components analysis were also used to compare the pattern of correlations, as well as the magnitude of the correlations among variables and factors in the data sets.

For all analyses, including correlation and principal components analysis, "missing data" (instances in which individual students had not responded to particular items) were handled through the statistical procedure of "listwise deletion" (also known as "casewise deletion"). This procedure deletes from the specific analysis the responses to all variables by the individual students with "missing data" (Nicholson 1989). By using "listwise deletion", a case or observation is eliminated if it has a missing value on any variable in the list. If many cases have missing data for some variables, listwise missing value treatment could eliminate too many cases and leave only a very small sample (SPSS, B-147). The other alternative, "pairwise deletion", deletes from the specific analysis only the missing data points, but is less appropriate because it may prevent estimation of coefficients due to singularities, may cause anomalies such as negative eigenvalues in factor analysis, and make hypothesis tests optimistic (Wilkinson, 1986).

CSE Subscales and Primary/Secondary Trait Scales

Item by item internal reliabilities for the four CSE Subscales and the Primary and Secondary Trait rubrics were conducted, and it was decided that no items would be deleted. Cronbach's Alpha was calculated for each subscale and trait scale as a measure of the reliability of scores for each measure. Table 5.2 summarizes these reliabilities for all groups included in both phases of the study. As is evident from this table, all alphas are high (above .80) with the exception of the Phase I Pretests alphas which were .75 and .64. These low alphas may be due to the fact that Phase I Pretest Writing Samples were difficult to gauge, especially by the two analytical subscales (CSE 4 and CSE 6). This gauging was difficult since specific decisions had to be made by each individual rater regarding the degree of errors (i.e., spelling and its effect on content).

Table 5.2

CSE Subscales and Trait Scales Pretest Internal Reliabilities

<u>Phase I</u>	<u>N</u>	<u>Mean</u>	<u>SD</u>	<u>Alpha</u>	<u>Items</u>
Control	20	17.9	3.0	.75	6
Treatment	20	16.0	2.6	.64	6
<u>Phase II</u>	<u>N</u>	<u>Mean</u>	<u>SD</u>	<u>Alpha</u>	<u>Items</u>
Treatment	26	20.4	5.9	.94	6

CSE Subscales and Trait Scales Post test Internal Reliabilities

<u>Phase I</u>	<u>N</u>	<u>Mean</u>	<u>SD</u>	<u>Alpha</u>	<u>Items</u>
Control	21	20.9	4.0	.83	6
Treatment	22	14.0	3.3	.86	6
<u>Phase II</u>	<u>N</u>	<u>Mean</u>	<u>SD</u>	<u>Alpha</u>	<u>Items</u>
Treatment	22	23.5	4.9	.87	6

Formative Tests Phase I: Formative Tests Item Reliabilities

FT1 and FT2 for the control and treatment groups in Phase I were subjected to item deletion procedures in order to test maximum

reliability. The lowest acceptable Alpha level was .50. In order to raise Alpha levels, different items were deleted for each different test and each group. Thus reliabilities for each Formative Test were done separately for each group. Item analysis was conducted, and item deletion was determined by item-total test score correlation and by the relative level of Alpha that inclusion of the item would yield in calculating the overall reliability of each test. Consequently, the best estimate of each test's reliability was based on different items for each group. For instance, the reliability of Formative Test 1, Part A, for the control group was best estimated by scores on 10 items out of a total of 12, as was the reliability for the treatment group. But the two items deleted in the treatment were not the same items as the ones deleted for the control group.

It was at times difficult to run certain analyses because of too few items, missing data points, and low n's. Still, after item deletion, all parts of Formative Test 1 Phase I had moderate levels of internal reliabilities for both the control and treatment class. These reliabilities ranged from .64 to .76. Only Formative Test 1, Part B, for the control group had a very low reliability of .03. Since there were few items in this section and the combined Alpha levels of part A and B yielded a .68, it was considered an adequate level of reliability. All Alpha levels for the treatment group were adequate, thus no further modifications or deletions were made.

Tables 5.3 and Table 5.4 present Formative Test 1 and 2 internal reliabilities. At times, it was necessary to combine two parts of a test in order to have enough items to establish a minimum Alpha level. For example, if Part B of Formative Test 1 could be eliminated completely, the items could and were included in combination with other parts of the test to get an Alpha level that was more representative of an instrument that might only include "good" reliable items.

Table 5.3

Phase I Formative Test 1 Internal Reliabilities
(after item deletion procedures)*
Formative Test 1

Part A	Mean	SD	Alpha	N	Items	Items Deleted
Control	6.4	2.4	.76	21	10	A1,A6
Treatment	4.8	2.4	.76	22	10	A7,A10
Part B	Mean	SD	Alpha	N	Items	Items Deleted
Control	6.6	.57	.03	21	3	B1,B2,B6,B7
Treatment	2.7	1.2	.64	22	4	B3,B6,B7
Part A,B	Mean	SD	Alpha	N	Items	Items Deleted
Control	13.0	2.4	.68	21	13	same as above
Treatment	7.6	2.6	.68	22	14	same as above

Table 5.4

Phase I Formative Test 2 Internal Reliabilities
(after item deletion procedures)*
Formative Test 2

Part A	Mean	SD	Alpha	N	Items	Items Deleted
Control	4.6	.74	.57	24	5	A5
Treatment	3.0	1.2	.73	24	4	A2,A3
Part B	Mean	SD	Alpha	N	Items	Items Deleted
Control	6.2	1.1	.65	24	7	B4
Treatment	5.2	1.5	.71	24	7	B8
Part C	Mean	SD	Alpha	N	Items	Items Deleted
Control	-	-	-	-	4	-
Treatment	-	-	-	-	3	C1
Parts ABC	Mean	SD	Alpha	N	Items	Items Deleted
Control	14.5	1.5	.61	24	16	same as above
Treatment	10.7	2.8	.81	24	14	same as above

Note that a low number of items prevented any separate reliability analysis for Formative Test 2, Part C, but these items were included in the overall reliability analysis. The control group had 4 items in part C, and the treatment group had only 3 items in part C.

Summative Test Reliabilities Phase I

Since the Summative Test consisted of four items only, it was more difficult to consider deleting items in order to raise the level of

Alpha; therefore no items were deleted. As evident from Table 5.5, reliability was low in both the control and treatment groups.

Table 5.5

Phase 1 Internal Reliabilities of Summative Test

	Mean	SD	Alpha	N	Items
Control	3.7	.61	.47	24	4
Treatment	2.6	1.0	.50	23	4

Summative Test Reliabilities Phase II

All items for all Summative Tests in Phase II were included in conducting procedures for establishing reliability. Since there were few items and the N was low (19), it was too difficult to delete any items in order to raise the Alpha level. Correlation coefficients also tend to be lower when sample sizes are small, such as in this Phase. These results were inevitable. The results of the Summative tests reliabilities are presented by Tables 5.6.

Table 5.6

Internal Reliabilities of Summative Tests Phase II

Mean	SD	Alpha	N	Items
8.4	.81	.35	19	9

Reliabilities of the Affect Questionnaire: Perception of Achievement, Attitude and Interest Questionnaire

Phase I

Estimating maximum internal reliabilities for the items in the various affective questionnaires was highly problematic. The major constraints were in identifying the items that were to be discarded to raise the level of Alpha. Each set of items for each part of the questionnaire were examined for each group separately. As low reliability items were discarded, an extremely disparate pattern was emerging--each administration of the questionnaire and each section

had different items that resulted in low reliability. It was difficult to decide what items to retain or discard for the final instrument assessment.

Too many missing data points and low Ns prevented the possibility of any further empirical analyses. As a result, the analysis became exploratory in nature. It consisted primarily of an item by item substantive examination of reliabilities.

Tables 5.7, 5.8, and 5.9 present the "best guess" of the estimated internal reliabilities for all three questionnaires for the control and treatment in Phase I. The Alpha coefficients range from a low of .32 to a high of .84. The highest levels of Alpha for all questionnaires occurred in Part C, the Interest dimension of the questionnaire. Part A had the next highest; Part B generally had the lowest. The reliabilities for the treatment were lower than for the control group. Table 5.9, the third administration of the questionnaire, had somewhat higher Alphas even though no items were deleted for any sections because of the low Ns for each group.

Table 5.7

Phase I Internal Reliabilities of Affect Scales:
Questionnaire I Affect Scale 1

<u>Part A</u>	<u>Mean</u>	<u>SD</u>	<u>Alpha</u>	<u>N</u>	<u>Items</u>	<u>Items Deleted</u>
Control	16.	3.0	.75	23	7	A1,A2,A5,A6,A11,A13
Treatment	31.0	2.2	.47	9	12	A8
<u>Part B</u>	<u>Mean</u>	<u>SD</u>	<u>Alpha</u>	<u>N</u>	<u>Items</u>	<u>Items Deleted</u>
Control	9.2	2.1	.51	23	4	B1,B4,B5
Treatment	15.8	3.0	.63	16	7	None
<u>Part C</u>	<u>Mean</u>	<u>SD</u>	<u>Alpha</u>	<u>N</u>	<u>Items</u>	<u>Items Deleted</u>
Control	14.2	3.4	.84	23	6	None
Treatment	15.1	2.6	.66	19	6	None

Table 5.8

Phase I Internal Reliabilities of Affect Scales
Questionnaire II Affect Scale II

Part A	Mean	SD	Alpha	N	Items	Items Deleted
Control	16.0	2.8	.65	24	7	A1,A2,A6,A8,A11,A13
Treatment	31.5	3.3	.50	18	13	None
Part B	Mean	SD	Alpha	N	Items	Items Deleted
Control	16.9	2.6	.54	24	7	None
Treatment	15.8	2.3	.32	21	7	None
Part C	Mean	SD	Alpha	N	Items	Items Deleted
Control	13.7	3.1	.82	25	6	None
Treatment	15.8	2.5	.70	21	6	None

Table 5.9

Phase I Internal Reliabilities of Affect Scales
Questionnaire III Affect Scale III

Part A	Mean	SD	Alpha	N	Items	Items Deleted
Control	30.5	4.7	.82	15	13	None
Treatment	31.5	3.9	.75	12	13	None
Part B	Mean	SD	Alpha	N	Items	Items Deleted
Control	16.6	2.9	.59	17	7	None
Treatment	15.8	2.8	.66	14	7	None
Part C	Mean	SD	Alpha	N	Items	Items Deleted
Control	12.8	3.7	.81	16	6	None
Treatment	15.7	2.1	.55	19	6	None

Phase II

Table 6.0 shows the internal reliabilities of all three administrations of the same questionnaire during Phase II. A cursory examination of the internal reliabilities for the Attitude/Perception Questionnaire reveals more consistency in the responses than in Phase I, but basically the Alpha levels remain low. Even though the n's were higher and more consistent across sections, and different administrations of the questionnaire, deleting items did not necessarily

raise the Alpha levels. The section that had the highest reliabilities was again section C for all the questionnaires.

Table 6.0

Phase II Internal Reliabilities of Affect Scales
Questionnaires I,II and III Affect Scale 1,II and III

<u>Questionnaire I Part</u>	<u>Mean</u>	<u>sd</u>	<u>Alpha</u>	<u>N</u>	<u>Items</u>
A	31.1	3.8	.66	26	13
B	16.5	3.1	.62	26	7
C	15.6	4.3	.87	29	6
<u>Questionnaire II Part</u>	<u>Mean</u>	<u>sd</u>	<u>Alpha</u>	<u>N</u>	<u>Items</u>
A	31.1	3.7	.68	26	13
B	16.1	2.8	.44	26	7
C	13.7	3.3	.75	26	6
<u>Questionnaire III Part</u>	<u>Mean</u>	<u>sd</u>	<u>Alpha</u>	<u>N</u>	<u>Items</u>
A	31.1	3.8	.66	26	13
B	15.8	2.5	.36	28	7
C	15.0	3.0	.80	25	6

Summary of Instrument Reliability

Tables 6.1 summarizes the reliability of all measures used in the study for both Phase I and Phase II.

Table 6.1

Summary of Instrument Reliability

	Phase I	Phase II
CSE Subscales	+	+
Trait Scales	+	+
Formative Tests	?	
Summative Tests	?	?
Affect Questionnaire	?	?

As has been seen in the previous section on reliability, the pre and post achievement measures tend to have the greater reliability than the affect ones. The Formative and Summative Tests in Phase I vary among the different sections of each test. Thus, it can be surmised that the reliabilities of these measures are questionable. Similarly, the reliability of the Summative Tests in Phase II (recall that only Summative scores were used for final gauging of achievement in this phase) is rather low

(.35). Only Part C of the affect questionnaire had acceptable reliabilities for each questionnaire administrations; Parts A and B did not.

Construct Validity of Measures

According to Baker (1982) relatively little reflection is given to the assessment part of writing instruction in practice. In addition, the criteria used to judge or rate the writing itself is usually unspecified. This was not the case for this particular study. The specification of the instructional task was based on clear identification of the characteristics of the task. The intent was not to reduce or fragment, but rather to delineate the specific skills being assessed. The initial step was the creation of specifications which were used to delimit the achievement domain of the writing task. The rules in a set of specifications circumscribe the content and skill to be measured and theoretically, an infinite set of test items might be constructed to measure the specified domain, which is a definitionally generalizable task. The domains in this study are identified in the Table of Specification as Instructional Units.

A basic tenet of criterion referenced measurement is that instructional treatments can be designed to teach the specified tasks. Thus, according to this premise, writing is a skill thought to be amenable to instruction. The task specification spells out what the critical features of learning to write. Therefore in writing tasks, the specifications permits, first, the identification of focused assessment and, secondly, the design of instruction to meet the critical features. Moreover, according to Baker (1982), to measure higher order skills such as essay writing, creativity, and other representations of complex human achievement, no clear, widely accepted definitions exist. All of the tests used for the assessment and gauging of writing skills, however, were designed based only on the instructional program used for the study. The items presented in the test formats were conceptually

similar to the ones used for instruction. Hence, these measurements can be considered valid only for measuring within-program achievement.

In summary, the construct validity of a test is the extent to which you can be sure it represents the construct of interest to actual classroom instruction. In this case, the construct was students' writing achievement. This construct was tested here by criterion-referenced tests, matched to objectives for the design and sequencing of the instruction. The template for these matches were the Table of Specifications for each instructional unit which specifically defined the parameters for diagnostic measurement.

The construct validity of the measures used in this study was tested in two ways: correlational methods which were used for the pre and post measures and principle components analysis for the Formative and Summative Tests and for the Attitude/Perception questionnaire. Correlational methods for supporting test validity are not often used with criterion-referenced tests because the scores from these tests may not have sufficient variability to produce reasonable correlation coefficients. However, correlational methods can assist in assessing construct validity because they help in identifying items that are inadequate for measuring the construct. Generally discarding those items will improve the quality of the instrument.

Since the achievement construct in question is a complex one, short separate tests to measure one or two subcomponents at a time were used. Thus a whole battery of tests addressed the principle construct. These separate tests varied in length, scope, and focus. There were the Formative and Summative Tests used in conjunction with the instruction and the scoring rubrics of the selected and modified CSE Subscales and Trait Scales used as Pre and Posttests.

The specifications for assessment of paragraph writing skill and instructional objectives for this study are as closely related as possible. There was a serious attempt to define the simplest subdivision of writing

tasks which dealt with what was to be used to "prompt " the writing act, what process, if any, the student would use in writing, and what criteria would be applied to judge the quality of effort. The topics or themes used to prompt the writing act were promoted consistently (in both phases of the study, although they were different for each phase) during the 'pre-writing' experiences. Thus, it could be argued that the content base was not inhibited by content restriction but rather attempted to draw on the students' real life as well as classroom experiences. The use of writing stimuli around the same 'concept' was deliberate and consistent but left generally unorganized (there was more organization of ideas for the Phase II writing assignment than for Phase I) so that the individual student could use experientially stored knowledge (see Chapter III and Chapter IV for specific examples for each phase of the study).

As we shall see, especially in the Formative and Summative Tests, correlational methods were useful to a point in this study since each subcomponent of each test had just a few items. The inter-correlations of the CSE Subscales and the CSE Subscales with the Trait Scales show evidence of the validity of a multiple measurement approach.

Construct Validity of CSE Subscales and Trait Scales

For lack of a better method for assessment and also to more closely match this study's instructional objectives, the wording of the original CSE Subscales was changed. This modification changed the focus of the testing from essay to paragraph. When these changes to the CSE Subscales were considered, several professionals identified as experts in the teaching of writing and composition research were consulted. These experts felt that the units of an essay were 'thought' units that should be components of a paragraph.

In actuality, few words of the original selected subscales were changed. Two of the subscales from the original CSE Expository Scale

IV were not used for this study, CSE Subscale 3--Essay Coherence-- and CSE Subscale 5--Support-- which deals with support features throughout an essay. It was felt by the researcher that the rating categories of these two subscales were not amenable to change for the purpose of assessing a single paragraph writing sample. Thus, two holistic subscales--CSE 1, General Impressions and CSE 2, General Competence--and two analytical--CSE 4, Paragraph Coherence and CSE 6, Mechanics had features that were useful for the assessment focus of this study.

Both sets of measures used for the assessment of the paragraph writing samples, the modified CSE Subscales and the Trait Scales, presume to assess expository writing. To test the construct validity of these modified measures and trait scales the results of each separate measure were correlated by using Pearson's Correlation Coefficient, listwise deletion procedures to see to what extent each measure would correlate with another thought to assess the same or a similar construct.

Findings in Table 6.2 report these correlations. As expected, they show that there was evidence of moderate positive correlations among measures. For instance, CSE 4 Paragraph Coherence, where the rater is asked to judge overall competence based on specific elements such as main idea, organization, coherence, support, and mechanics, was positively correlated (Posttest reliabilities were .80, .82) with CSE 2 General Competence which evaluates the paragraph as a 'whole.' The Posttests correlation coefficients between CSE 4 Paragraph Coherence and CSE 1 General Impressions were also strong, ranging from .82 to .87. Similarly CSE 1 correlated strongly with CSE 2 (with the exception of Posttest R2) suggesting that there was considerable overlap in the raters perception of competence and their first impressions of the paragraph samples produced by students.

Table 6.2

Intercorrelations of CSE Subscales
Pretest Intercorrelations

	CSE 1*		CSE 2		CSE 4	
	R1	R2	R1	R2	R1	R2
CSE2 General Competence	.80	.85	--	--	--	--
CSE4 Paragraph Coherence	.64	.75	.72	.72	--	--
CSE6 Mechanics	.58	.44	.61	.33	.55	.40

*CSE 1 General Impressions

Posttest Intercorrelations

	CSE 1*		CSE 2		CSE 4	
	R1	R2	R1	R2	R1	R2
CSE2 General Competence	.84	.63	--	--	--	--
CSE4 Paragraph Coherence	.80	.87	.84	.82	--	--
CSE6 Mechanics	.62	.76	.73	.76	.68	.70

*CSE1 General Impressions

The construct presumed to be measured by CSE 6 Subscale Mechanics has lower correlations with the other subscales especially at Pretest. That this domain may not necessarily be related to composition was substantiated by Polin (1985). Polin, whose study proposed to test the construct validity of current assessment methods by providing activities designed to assist, not instruct, students in planning and revising behaviors described as requisites of good writing, also identified the CSE 6 Mechanics subscale as being independent from the focus of the other subscales. She explains this independence by stating that "grammar and punctuation skills which many educators consider to be part of a separate domain of English fluency, not necessarily related to composition. Thus, the Mechanics subscale would not be expected to share quite a strong relationship with other subscales."(p.102)

While the CSE Subscale 6 Mechanics at Pretest level appears to be independent from all the other subscales yielding coefficients ranging from as low as .33 to a low moderate of .61, by Posttest there appears to be a shift. This shift is evidenced by the Posttest correlation coefficients which show a narrower range from a low of .62 to a moderate .76 among the various scales. This change may indicate an influence of the

instructional objectives, especially for Phase I of the study. A lot of emphasis was placed on punctuation, capitalization, and indentation in the instructional activities and on the Formative and Summative Tests used in this phase. Remember that the focus of the classroom activities, unlike Polin's study, was to instruct for the purpose of learning not just to assist the learning process.

As was expected the constructs presumably assessed by the two "holistic" Trait Scales, the Primary Trait "Main Idea" scale and the Secondary Trait "Inventive Expression and Elaboration on a topic" scale appear to be independent. Table 6.5 presents the correlations among the two trait scales. Recall that the trait scales were developed specifically to "holistically" assess paragraph main idea in the Primary Trait and Inventive Expression and Elaboration on the topic in the Secondary Trait. Also recall that the Secondary Trait, unlike the CSE or the Primary Trait, was developed basically to take into account the creative language production of students. If traits are construct valid, then we would expect that they not correlate highly with each other (as in Table 6.3), that there would be low correlations for the Secondary trait with the CSE Subscales, and that there would be high Primary trait correlations with the CSE Subscales. As illustrated by Table 6.5, the magnitude of the intercorrelations among these two scales is similar to those between the CSE and the Trait Scales (refer to Table 6.3). The difference, however, between these comparisons is that these scales, unlike the CSE Subscales, were scored along the same point range. They both had five categories or levels for assessment. The Posttest coefficients are only slightly higher than the Pretest coefficients ranging from .58 to .61 among the two raters. For the most part, all the coefficients appear to be low enough to warrant the assumption that both scales are gauging sufficiently different constructs. Although theoretically, and as stated in the Secondary Trait description, the best writing samples are highly structured and are coherent, structure does not define content, but rather content defines structure. Again as with the intercorrelations

among the CSE Subscales and the Trait scales, all skills considered together provide a comprehensive dimension of the construct labeled paragraph writing skill.

Table 6.3

Intercorrelations: Primary Trait Main Idea Scale
with Secondary Trait Inventive Elaboration Scale

PRETESTS

Secondary Trait	Primary Trait	
	Primary R1	Primary R2
Secondary R1	.56	.57
Secondary R2	.55	.56

POSTTESTS

Secondary Trait	Primary Trait	
	Primary R1	Primary R2
Secondary R1	.61	.58
Secondary R2	.60	.58

Campbell and Fiske (1959) proposed that in order to demonstrate construct validity, it is necessary to show not only that a test correlates highly with other variables with which it should correlate, but also that it does not correlate with variables from which it should differ.

Anastasi (1964) states that essentially what is required is the assessment of two or more traits by two or more methods. The correlations of the same trait assessed by different methods represent a measure of convergent validity (these should be high). The correlations of different traits assessed by the same or similar methods provide a measures of discriminant validity (these should be low or negligible). To some extent, the construct validity by the intercorrelations of the various methods used in this study, is substantiated.

Contrary to this assumption, Table 6.4 indicates low moderate intercorrelations for all the Posttests ranging from .62 to .71. Note especially that at Pretest level CSE 4, Paragraph Coherence, have very low intercorrelation coefficients with the Secondary Trait ranging from .42 and .49, but that by Posttest level these correlations increase to .63 for both raters. The only exception appears to be the CSE 6, Mechanics

subscale. These correlations seem to indicate that all skills when considered together produce higher levels of inventive expression and elaborative language and that (at least with the intercorrelation with this scale) Mechanics can be seen as separate from the constructs assessed by the other scales. The highest CSE 6 Mechanics intercorrelation coefficient with Secondary Trait at Pretest is .48 which only increases to .50 by Posttest level.

Table 6.4

Intercorrelations: CSE Subscales with Primary and Secondary
Pretests

	Primary		Secondary	
	R1	R2	R1	R2
CSE 1 General Impression	.50	.50	.66	.63
CSE 2 General Competence	.59	.56	.61	.66
CSE 4 Paragraph Coherence	.55	.56	.42	.49
CSE 6 Mechanics	.32	.20	.48	.40

Posttests

	Primary		Secondary	
	R1	R2	R1	R2
CSE 1 General Impression	.61	.59	.69	.71
CSE 2 General Competence	.50	.57	.62	.64
CSE 4 Paragraph Coherence	.57	.52	.63	.63
CSE 6 Mechanics	.46	.51	.50	.51

Construct Validity: Formative and Summative Tests

Principal components analysis analyzes the total variance of the items (or variables). It is technically a transformation of the data into a set of orthogonal variables rather than a factor analysis (Tinsey and Tinsey, 1987). Thus a factor loading (magnitude of the item) represents the correlation of the variable (item) with the factor. As already mentioned, principal components analysis was used as a descriptive exploratory procedure which only with replication on new samples of subjects or variables can serve to determine the extent to which the same factors tend to emerge.

Principal components solutions to test construct validity of the Formative and Summative Tests were generated by using Varimax rotations, but a problem arose. The intent had been to generate a

principal components test using most items after reliability execution of Cronbach's Alpha procedures. After using Cronbach's Alpha correlation methods to estimate maximum test reliabilities, it proved impossible to run a principal components analysis. Most combinations of principal components solutions would not run because of an insufficient number of items per test. For this reason, the decision was made to run these analyses by including the originally discarded item. Thus, the following principal components tables should be taken as tainted evidence of construct validity for the Formative and Summative Tests of both phases of the study.

Table 6.5 represents the underlying constructs of six factor solution by the Formative and Summative Tests given as regular quizzes in the control classroom. Systat routines indicated the primary factor loadings; thus items are listed and presented in order as sorted. The cut-off point was .50; therefore items below this level were not reported since they indicate only modest to low correlations.

Table 6.5

Phase I-Principal Components of 6 Factors Formative Tests 1 and 2 Control Group

Items	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
FT1U1A4	.935 wrd.mean					
FT2U2A6	.924 punct.					
FT1U1A3	.833 wrd.mean					
FT1U1A5	.799 wrd.mean					
FT2U2B4*	.587 punct.					
FT2U2A2	.547 punct.					
FT1U1A2	.526 wrd.mean					
FT2U2A4		-.971 punct.				
FT2U2B5		-.971 cap.				
FT2U2B6		-.971 punct.				
FT2U2B7		-.686 cap.				
FT2U2A1		-.680 punct.				
FT2U2B8*		-.557 punct.				
FT1U1B1*			-.963 sent.mean			
FT2U2C3			-.963 cap.			
FT1U1B7*			-.818 sent.mean			
FT1U1A1*				-.824 wrd.mean		
FT1U1A6*				-.701 wrd.mean		
FT2U2B2				.610 cap.		
FT1U1A12				.514 wrd.mean		
FT1U1A9				-.512 wrd.mean		
FT2U2A5					-.689 punct.	
FT1U1A10					.606 wrd.mean	
FT2U2B3					.596 cap.	
FT1U1A8					.527 wrd..mean	
FT1U1A11						-.695 wrd.mean
FT1U1B3						-.672 sent.mean

**Note *items deleted to maximize reliabilites

The first factor consists of items from both Formative Test 1 and 2. It appears that the most salient underlying concept for this factor is word meaning. The items that deal primarily with punctuation (FT2U2A6, FT2U2A2, FT2U2B4) require the student to assess meaning in order to provide the proper 'end' punctuation to a sentence. All of the remainder of the items on this factor appear on the first Formative Test and deal with word meaning. Students were asked to circle 12 words from a list of 20 that could be used to describe how they felt or how things looked, felt, smelled, tasted, or sounded. The first unit and content objective delineated in the Table of Specifications was specifically the identification of abstract words. When these words were generated and defined they were used to describe abstract ideas about basic senses or emotions. In turn, these words were used to develop sentences. This preliminary unit sets the direction and the basis for the four week instructional sequence.

The second factor, which appears to be more homogeneous than the first factor, consists of all items with the exception of FT1U1B8 from Formative Test 2. Unit 2 in the Table of Specifications deals with grouping sentences, punctuation, and capital letters. This factor might best be labelled period/punctuation mark. The objectives were to recognize and use capital letters and identify the period as a punctuation mark. Two of the items in this cluster deal with beginning capitalization, but at least one of these items pairs off with the end punctuation of the same sentence that requires the capitalization (refer to Chapter III for specific examples of these items).

Factor Three is comprised of three items. Two of these items appear in the B Section of Test 1 and one item is from Section C, Test 2. This factor is best interpreted as sentence meaning as opposed to word meaning and punctuation or capitalization alone. The items that appear in this cluster load negatively. The item from Formative Test 2 that deals primarily with capitalization (FT2U2C3) requires the respondent to rewrite a sentence and provide the correct capitalization

and punctuation. Comprehension of sentence meaning is necessary in order to respond correctly. The meaning of the sentence which constitutes this item is ambiguous. It is a question that requires the respondent to use speech inflection in order to assess meaning. In other words, this is not just simple punctuation for a fourth grade student, especially if English is the second language. Punctuation in isolation from meaning requires automaticity and familiarity with the inflectional nuances of the language. Moreover, this particular item (FT2U2C3) contains the word 'does' in the form of a question; the transformation of the auxiliary verb 'do' is not a structure in the Spanish language. Perhaps these linguistic elements and the negative loadings on this items reveal some of the complex underlying cognitive demands inherent in nuances of language meaning and, specifically, sentence meaning.

It is more difficult to interpret Factor Four in this principal components solution than any other factor. The items that clustered on this factor have both negative and positive loadings. Only one item is from Formative Test 2 (FT2U2B2); all the other items are from Part A of Formative Test 1. All items from Part A Formative Test 1 deal with the underlying meaning of the verbal noun "to feel." The ambiguity of meaning that appears to be related to the notion of "to feel" has two interpretations. One meaning of "to feel" is related to emotions. The other meaning is "to feel" as tactile or physiological. Obviously, these two meanings are related concepts. The expectation though, of a specific level of language performance (the automaticity of response expected from native language fluency), versus language competence (may know the correct answer but require a longer time to process and respond), could have affected the individual responses of students with items such as these. In the three initial instructional units identified in the Table of Specification, the objectives included capitalization and punctuation--which required in an interspersed manner, recognition, identification--and application of these tools to get to word and sentence

meaning. Hierarchical complexity of meaning in order to "conceptually-piece language" to produce elements of a basic paragraph structure was the objective of the fourth and final instructional unit.

Even less clear are the loadings on Factor Five. This cluster includes one item that loads negatively (FT1U1A10) and three that load positively. The two items that emerge from Formative Test I (FT1U1A10, FT1U1A8) are from the same section but have unrelated content meanings. The two items from Formative Test 2 are not conceptually related either; one requires end punctuation and the other requires capitalization. This factor was considered uninterpretable.

Similarly, Factor Six was uninterpretable. It contains only two items. Both items are from the same test but different sections and are conceptually related; one requires word meaning and the other requires sentence meaning. The loading of only two items is not enough to generate a label for this factor.

Table 6.6 illustrates the factor loadings for the treatment classroom in Phase I of the study. Unlike the factor loading for the control classroom only one item that was deleted to maximize reliability appears to have loadings on the second factor (FT1U1A1). There are only three factors that loaded for this group of scores. Also, all factors include items that assess word meaning and sentence meaning. The third factor is homogeneous reflecting word meaning only. Finally, unlike the principal components analysis for the control group, one item (SUM3A3U3) from the Summative Test is included on the first factor loadings.

Table 6.6

Table 6.6

Phase I

Principal Components of 3 Factors

For Formative Tests and Summative Test Treatment Group

Items	Factor 1	Factor 2	Factor 3
FT2U2B5	.909punct.		
FT2U2A2	.872punct.		
FT2U2C4	.870punct.		
FT2U2B7	.821punct.		
FT1U1A5	.789wrđ.mean		
FT2U2B6	.723punct.		
SUM3U3A3	.663sent.order		
FT2U2C3	.631cap.		
FT1U1A3	.608wrđ.mean		
FT1U1A1	.585wrđ.mean		
FT2U2C2	.576punct.		
FT1U1A6	.553wrđ.mean		
FT2U2A6		.836punct.	
FT2U2A1		.771punct.	
FT2U2A5		.755punct.	
FT1U1A10*		.751wrđ.mean	
FT1U1B5		.703sent.mean	
FT2U2A4		.595punct.	
FT1U1B2		.541sent.mean	
FT1U1B1		.523sent.mean	
FT2U2B1		.541cap.	
FT1U1A8			.737wrđ.mean
FT1U1A11			.687wrđ.mean
FT1U1A4			.672wrđ.mean
FT1U1A9			.613wrđ.mean
FT1U1A12			.523wrđ.mean

What appears as the strongest factor in the principal components solution for the treatment group is Sentence and Word Meaning and their relation to capitalization and punctuation. The items that appear on this loading from Formative Test 1 represent all concepts related to the basic senses. This is conceptually similar to the main objective of the unit 1 instructional sequence. In addition, Factor One of the treatment appears to contain items that require the respondent to be familiar with nuances of English language inflection patterns (just as in Factor Three for the control group). These items also reveal that respondents must know how these inflection patterns relate to sentence meaning and the

appropriate capitalization and punctuation. For instance, item FT2U2C3 and item FT2U2C4 are conceptually related; one requires initial sentence capitalization and the other requires end punctuation. The content of the sentence however requires that the respondent have some familiarity with the inflectional patterns of English in order to determine that the sentence is a question. In addition the word "does" is the initial word in the phrase *does Joe have a dog*. This may cause confusion for students not familiar with the use of the complex structure with the form of the auxiliary verb *do* which requires this verb to be transformed into a grammatically acceptable question. As previously mentioned, the complex transformations of this auxiliary verb does not appear in Spanish language syntax.

The First Factor also loads an item from the Summative Test which requires sequencing sentences to generate a cohesive paragraph according to meaning. The sentences included distractor sentences that had no related meaning to the ones that were to be sequenced. The inclusion of this item on this factor loading supports the notion that sentence meaning and nuances of language was a skill that was being consistently assessed by these tests in combination.

The Second Factor dealt specifically with punctuation. The items that loaded on this factor appear to focus (with the exception of the item starred for deletion) on sentence meaning which required understanding of the inflection pattern related to the appropriate punctuation. For instance, FT2U2A6, FT2U2A1 and FT2U2A5 required complex punctuation for the different inflections relating to exclamation, question, and command sentence structures. The other items dealt with sentences that were declarative sentences. Thus all inflection modes were represented in the loading for this factor.

The Third and final Factor in this principal components solution for the treatment group focuses on word meaning. All of the items that loaded on this factor are conceptually similar and included items from section A of Formative Test I. The five items included in this cluster

deal with how things look. This factor appears to represent the most concrete or pragmatic concept with regard to the notion of the "basic senses" and word meaning.

Table 6.7 shows the results of a principal components analysis for all the Summative Tests used in Phase II of the study. Since these tests were cumulative in nature there were few items. No items were deleted to estimate maximum reliability (since Alpha was low and deleting items would not raise it anyway); the loading reflects cumulative concepts such as 'main idea.' Notice that one item reflects one factor for this group. These results simply reflect the nature of the Summative Tests.

Table 6.7

Phase II Treatment Principal Components with 4 Factors of Summative Tests

Item	Factor 1	Factor 2	Factor 3	Factor 4
SUMM1U1A1	-.975main idea			
SUMM1U1A2	-.975main idea			
SUMM1U1A3	-.825main idea			
SUMM2U2A2		.970 indent/main idea		
SUMM3U3A3			.996 sequence/main idea	
SUMM2U2A3				.956 indent/main idea

The loadings of items for the Phase II treatment principal components solutions were very difficult to interpret. While they appear homogeneous, the clusters do not necessarily reveal underlying structure. Since the variables presumably gauged by these items are similar with regard to complexity levels, the loadings may be a "catch-all" for each other in one factor but have little to do with underlying structures. Factor One for instance, loads all the items on the first Summative test. These items are correlated negatively. All three items are similar; they deal specifically with the main idea of a paragraph. Even though each item is actually a separate paragraph with distinct content, all items require the respondent to delete the phrase or sentence that does not belong in order to maintain coherency in content.

The other Three Factors each have loadings of one item each. Notice that Factor Two and Factor Four are from the same instrument

and instructional unit. They are in fact presumably gauging the same concept, first indenting, then sequencing three consecutive paragraphs on the same content idea. The Third Factor deals with sequencing sentences in a paragraph, crossing out the sentence that does not belong, and underlining the topic sentence. Too few items and not enough variability in responses made the assessment of underlying structures difficult for this phase of the study.

Construct Validity: Perception of Achievement,
Attitude and Interest Questionnaire

After considerable deliberation by the researcher it was concluded that the validity of the Perception of Achievement and Attitude questionnaire could not be defended. Although there was a concerted effort made to match the content of the items in the questionnaire to the instructional objectives in both phases of the study and also to provide an English, as well as a Spanish version, the principle components analysis which was tried several times with different combinations of items did not yield any usable results (refer to discussion of reliability).

The wording of the original items of the questionnaire was modelled after the items on several questionnaires used in Anaina's (1985) study (refer to Chapter III, procedures). The present study had a different content focus and was administered to students at one grade level as opposed to three different grade levels.

The inconsistency of results for the present study may be due to several factors. First, the procedures for administration and analysis that were used. A careful approach for administration of the questionnaire was followed but, given the problems of program implementation in Phase I of the study (refer to Chapter III), it seems likely that these affected the administration of the questionnaire, as well as the overall implementation of instructional objectives. Second, although the wording of the items followed the same format as Anaina's

study, the format of the overall questionnaire was different in that each part of the questionnaire (Part A, Part B, and Part C) focused on a slightly different dimension of affect (refer to Chapter III). In Aniana's study the three different components were actually different questionnaires with a slightly distinct dimensional focus. Another problem may be due to the students' lack of comprehension or self-awareness. A question which may appear to be understood by an adult in actuality, is not understood by the student. It is possible that the student may not be aware of his or her own attitudes. Young children may not be sufficiently introspective to know their own attitudes. These respondents were fourth grade students who perhaps had never filled out a questionnaire; especially one that required them to retrospectively gauge their own attitudes.

The third factor, which is possibly the most important in the present study and that may perhaps account for the discrepancies in results, is that the response pattern only required a response of 'yes', 'don't know', or 'no.' The 'don't know' response was a neutral response. Even though this was the same response pattern used in Aniana's study, there seem to be a bias in the response pattern since students only had three choices rather than the usual five choice Likert scale. The number of response options per question seems to have affected reliability and also the validity by restricting the range of possible scores. The content of the items in each section of the questionnaire included both positive and negative statements about the subject they were studying. Perception of achievement (the subjective judgments made by students about the level of learning they attain), was operationalized as students' responses to items taken from an instrument which required students to indicate how well they thought they were learning, to compare their work with the work of their classmates, and to project how their learning was viewed by others. Even so, it seems as if responses to the various items may be due to the desire of students to please the teacher rather than their true feelings.

Summary of the Construct Validity of Measures

Usually if the assessment instruments used in a study exhibit weak reliabilities, the validity of the measures is questionable. For this study this seems to hold true. Table 6.8 illustrates that the pre and post achievement measures had strong construct validity. This is true for both the CSE Subscales and Trait Scales. There appears to be high moderate correlations among the subscales and low moderate correlations between the subscales and the trait scales. The validity for the Formative Tests and the Summative Tests for both phases is questionable. Small sample sizes, missing data points, and insufficient variability in responses prevented adequate interpretation of underlying constructs. Similarly, the construct validity of the affective measure, the Perception of Achievement, Attitude and Interest Questionnaire was difficult if not impossible to defend.

Table 6.8

Summary of Instrument Validity

	Phase I	Phase II
CSE Subscales	+	+
Trait Scales	+	+
Formative Tests	?	
Summative Tests	?	?
Affect Questionnaire	?	?

CHAPTER VI

Findings

The treatment and independent variable manipulated was Mastery Learning. Each of the subjects in Phase I were measured at only one level of the independent variable, either conventional instruction or Mastery Learning. In order to test for student achievement, there were six dependent variables (four CSE Subscales and two Trait Scales) subjected to analyses. A repeated measures ANOVA was used.

Phase IBasic Data

In Phase I of this study it was expected that the Mastery Learning group would achieve at higher means than the control group on the Pre and Post measures. In addition, it was expected that the Mastery Learning group would have lower variance than the control group on all achievement indexes. Table 6.9. presents the means and standard deviations of the students' averaged scores on each measure for each subscale and each trait scale for Phase I.

Table 6.9

Phase I Overall Means and Standard Deviations: CSE and Trait Scales

	<u>Pretest</u>			<u>Posttest</u>		
	<u>n</u>	<u>Mean</u>	<u>sd</u>	<u>n</u>	<u>Mean</u>	<u>sd</u>
<u>CSE 1:General Impression</u>						
Control	29	3.1	.81	29	3.4	1.1
Treatment	25	3.1	.87	25	2.3	.73
<u>CSE 2:General Competence</u>						
Control	29	2.5	.59	29	3.5	.92
Treatment	25	2.9	.50	25	1.8	.45
<u>CSE 4:Paragraph Coherence</u>						
Control	29	2.9	.86	29	3.5	.96
Treatment	25	3.0	.63	25	2.1	.77
<u>CSE 6:Mechanics</u>						
Control	29	2.8	.80	29	3.7	.88
Treatment	25	2.4	.82	25	2.3	.82
<u>Primary Traits:Main Idea</u>						
Control	29	3.6	.73	29	3.7	.80
Treatment	25	3.1	.56	25	3.1	.89
<u>Secondary Trait: Inventive</u>						
Control	29	2.8	.85	29	2.7	.81
Treatment	25	2.6	.99	25	2.3	.70

As is evident in Table 6.9, the overall impression of subscale means is that the control and treatment groups are approximately equal on all Pretest CSE Subscales (only differing by .4). In other words, there does not seem to be a great difference in entry skill level between the groups. The variances at Pretest level also do not differ greatly (differ only by .9), with only one exception--the CSE 4 subscale where the treatment group variance is .23 lower than the control group. Posttest CSE means, however, are consistently higher for the control group than for the treatment group. The reverse is true for Posttest standard deviations, though.

The Primary and Secondary Trait Pretest means are also roughly equivalent (.5 and .2 difference respectively), as are the variances (.4 for the Primary Trait and .2 for the Secondary Trait). Both Posttest scores are greater for the control group than for the treatment group. The treatment group's Posttest variance is only somewhat less than the control's on the Primary Trait and on the Secondary Trait .

Data Analysis

Repeated measures ANOVA was conducted for each of the dependent variables testing the acquisition of writing skill in a single paragraph form. To reiterate, the subscales used to assess each dimension of writing skill is a dependent variable represented by CSE Subscales 1, 2, 4, 6 and two Trait Scales, Primary and Secondary. In this section each of these separate analyses will be presented.

CSE Subscale 1: General Impressions

Table 7.0 presents results of repeated measures analysis of variance for Phase I, CSE Subscale 1, General Impressions. These results shows that there was a significant interaction at $p \leq .05$. The comparison of the two means across repeated measures for Pre and Posttest (see Table 7.1) reveals that the teaching approach as well as the

difference of test administration (time) differentially affected the groups. The expectation was that students in the Mastery Learning classroom would outperform students in the control situation. As is evident in Table 7.1, the control classroom outperformed the students in the Mastery Learning classroom, with the control group means increasing over time while the treatment group means decreased.

Table 7.0

**REPEATED MEASURES ANOVA
CSE SUBSCALE 1: GENERAL IMPRESSIONS**

19 CASES DELETED DUE TO MISSING DATA

NUMBER OF CASES PROCESSED 35

UNIVARIATE AND MULTIVARIATE REPEATED MEASURES ANALYSIS

BETWEEN SUBJECTS EFFECTS

TEST FOR EFFECT CALLED: Teaching Method

SOURCE	SS	DF	MS	F	P
HYPOTHESIS	6.073	1	6.073	7.845	0.008*
ERROR	25.548	33	0.774		

WITHIN SUBJECTS EFFECTS

TEST FOR EFFECT CALLED: Time of Test

SOURCE	SS	DF	MS	F	P
HYPOTHESIS	1.122	1	1.122	1.191	0.283
ERROR	31.071	33	0.942		

TEST FOR EFFECT CALLED: Teaching Method X Time of Test

SOURCE	SS	DF	MS	F	P
HYPOTHESIS	4.979	1	4.979	5.288	0.028*
ERROR	31.071	33	0.942		

Table 7.1

**OVERALL MEANS DEPENDENT VARIABLE: CSE Subscale 1
General Impressions**

	PRPT	POPTS	DIFFERENCE
CONTROL N=16	3.1	3.4	+.3
TREATMENT N=19	3.1	2.3	-.8
DIFFERENCE	0	-1.1	

CSE Subscale 2: General Competence

Table 7.2 presents the results for the repeated measures ANOVA of CSE 2: General Competence. The analysis of variance found a

significant interaction effect at $p \leq .01$. As Table 7.3 indicates, at Pretest level both groups seem to be at comparable level of performance differing only by .4. By Posttest level, however, the treatment group means decrease by 1.8 and the control group means increase by as much as 1.0. These data indicate that one teaching approach (conventional) increased general competency from Pre to Posttest while the other teaching approach (mastery) decreased it.

Table 7.2

REPEATED MEASURES ANOVA
CSE SUBSCALE 2: GENERAL COMPETENCE

19 CASES DELETED DUE TO MISSING DATA

NUMBER OF CASES PROCESSED 35

UNIVARIATE AND MULTIVARIATE REPEATED MEASURES ANALYSIS

BETWEEN SUBJECTS EFFECT

TEST FOR EFFECT CALLED: Teaching Method

SOURCE	SS	DF	MS	F	P
HYPOTHESIS	15.241	1	15.241	36.900	0.000*
ERROR	13.630	33	0.413		

WITHIN SUBJECTS EFFECTS

TEST FOR EFFECT CALLED: Time of Test

SOURCE	SS	DF	MS	F	P
HYPOTHESIS	0.888	1	0.888	1.931	0.174
ERROR	15.183	33	0.460		

TEST FOR EFFECT CALLED: Teaching Method X Time of Test

SOURCE	SS	DF	MS	F	P
HYPOTHESIS	11.260	1	11.260	24.473	0.000*
ERROR	15.183	33	0.46		

Table 7.3

OVERALL MEANS DEPENDENT VARIABLE: CSE Subscale 2
General Competence

	PRPT	POPTS	DIFFERENCE
CONTROL N=16	2.5	3.5	+1.0
TREATMENT N=19	2.9	1.8	-1.6
DIFFERENCE	-.4	-1.7	

CSE Subscale 4: Paragraph Coherence

Table 7.4 presents results for repeated measures analysis of variance for the CSE Subscale 4 for Paragraph Coherence. Again a significant interaction effect ($p \leq .01$) was found. Table 7.5 indicates that the means of the control group stayed relatively constant from Pre to Post while and the means of the treatment group actually decreased over this time.

Table 7.4

REPEATED MEASURES ANOVA
CSE SUBSCALE 4: PARAGRAPH COHERENCE
19 CASES DELETED DUE TO MISSING DATA
NUMBER OF CASES PROCESSED 35
UNIVARIATE AND MULTIVARIATE REPEATED MEASURES ANALYSIS
BETWEEN SUBJECTS EFFECTS

TEST FOR EFFECTS CALLED: Teaching Method

SOURCE	SS	DF	MS	F	P
HYPOTHESIS	7.880	1	7.880	19.746	0.000*
ERROR	13.170	33	0.399		

*** WITHIN SUBJECTS EFFECTS***

TEST FOR EFFECTS CALLED: Time of Test

SOURCE	SS	DF	MS	F	P
HYPOTHESIS	0.214	1	0.214	0.204	0.654
ERROR	34.607	33	1.049		

TEST FOR EFFECT CALLED: Teaching Method X Time of Test

SOURCE	SS	DF	MS	F	P
HYPOTHESIS	8.628	1	8.628	8.228	0.007*
ERROR	34.607	33	1.049		

Table 7.5

OVERALL MEANS DEPENDENT VARIABLE : CSE Subscale 4
Paragraph Coherence

	PRPT	POPTS	DIFFERENCE
CONTROL N=16	2.9	3.5	+.6
TREATMENT N=19	3.0	2.1	-.9
DIFFERENCE	+.1	-1.4	

CSE Subscale 6: Mechanics

Table 7.6 presents the analysis of variance for CSE Subscale 6: Mechanics. A significant interaction effect at $p \leq .01$ appears once again. As with the other subscales, Table 7.7 shows a divergent trend in the direction of the means for both teaching approaches. The mean performance was highest at Pre test level and lowest at Posttest level for the Mastery group and the reverse for the control group. Thus, the control group means increase while the treatment group ones decrease.

Table 7.6

REPEATED MEASURES ANOVA					
CSE SUBSCALE 6: MECHANICS					
19 CASES DELETED DUE TO MISSING DATA					
NUMBER OF CASES PROCESSED 35					
UNIVARIATE AND MULTIVARIATE REPEATED MEASURES ANALYSIS					
BETWEEN SUBJECTS EFFECTS					
TEST FOR EFFECTS CALLED: Teaching Method					
SOURCE	SS	DF	MS	F	P
HYPOTHESIS	14.138	1	14.138	16.153	0.000*
ERROR	28.884	33	0.875		
WITHIN SUBJECTS EFFECTS					
TEST FOR EFFECT CALLED: Time of Test					
SOURCE	SS	DF	MS	F	P
HYPOTHESIS	2.432	1	2.432	5.240	0.029*
ERROR	15.318	33	0.464		
TEST FOR EFFECT CALLED: Teaching Method X Time of Test					
SOURCE	SS	DF	MS	F	P
HYPOTHESIS	4.918	1	4.918	10.595	0.003*
ERROR	15.318	33	0.464		

Table 7.7

OVERALL MEANS DEPENDENT VARIABLE: CSE Subscale 6
Mechanics

	PRPT	POPTS	DIFFERENCE
CONTROL N=16	2.8	3.7	+.9
TREATMENT N=19	2.4	2.3	-.1
DIFFERENCE	-.4	-1.4	

Summary of Repeated Measures
ANOVA: CSE Subscales

The results of the repeated measures analysis of variance for the CSE Subscales indicate that for all measures there were interactions. These interactions indicate that the groups were differentially affected by the treatment across time. The expectation was that the Mastery treatment group would outperform the control group by Posttest on all measures. This expectation was not realized. Although both groups appeared at roughly equivalent Pretest skill level, by Posttest the control scores increased overtime while the treatment group scores decreased.

Trait Scales: Primary and Secondary Phase I

While various dimensions of the CSE Subscales assess paragraph structure, the Primary Trait Scales was specifically developed to "holistically" assess main idea in a paragraph. Therefore, the trait is the presence or absence of an explicit main idea sentence in a given paragraph. The Secondary Trait assesses inventive elaboration of a topic.

Primary Trait: Main Idea

Table 7.8 presents the repeated measures of analysis of variance results for Primary Trait main idea indicate a significant main effect between the two classes ($p \leq .01$). This effect indicates the groups were significantly affected by the teaching approach. As noted in Table 7.9, the trait means for the control increased from Pre to Posttest and remained virtually the same for the treatment.

Table 7.8**REPEATED MEASURES ANOVA****PRIMARY TRAIT: Main Idea**

19 CASES DELETED DUE TO MISSING DATA

NUMBER OF CASES PROCESSED 35

UNIVARIATE AND MULTIVARIATE REPEATED MEASURES ANALYSIS

BETWEEN SUBJECTS EFFECTS**TEST FOR EFFECTS CALLED: Teaching Method**

SOURCE	SS	DF	MS	F	P
HYPOTHESIS	6.538	1	6.538	10.762	0.002*
ERROR	20.048	33	0.608		

WITHIN SUBJECTS EFFECTS**TEST FOR EFFECTS CALLED: Time**

SOURCE	SS	DF	MS	F	P
HYPOTHESIS	0.068	1	0.068	0.119	0.733
ERROR	18.875	33	0.572		

TEST FOR EFFECT CALLED: Teaching Method X Time

SOURCE	SS	DF	MS	F	P
HYPOTHESIS	0.068	1	0.068	0.119	0.733
ERROR	18.875	33	0.572		

Table 7.9**OVERALL MEANS DEPENDENT VARIABLE:****Primary Trait Scale**

	PRPT	POPTS	DIFFERENCE
CONTROL N=16	3.6	3.7	+.1
TREATMENT N=19	3.1	3.1	0
DIFFERENCE	-.5	-.6	

Secondary Trait: Inventive Elaboration of Topic

Table 8.0 repeated measures analysis of variance shows a significant main effect between the control and experimental groups ($p \leq .10$). As can be seen from Table 8.1, the scores for both groups from Pre to Posttest decrease. This decrease is sharper for the experimental group than for the control.

Table 8.0**REPEATED MEASURES ANOVA****SECONDARY TRAIT: Inventive Elaboration of Topic**

20 CASES DELETED DUE TO MISSING DATA

NUMBER OF CASES PROCESSED 35

UNIVARIATE AND MULTIVARIATE REPEATED MEASURES ANALYSIS

BETWEEN SUBJECTS EFFECTS**TEST FOR EFFECTS CALLED: Teaching Method**

SOURCE	SS	DF	MS	F	P
HYPOTHESIS	2.317	1	2.317	3.260	0.080*
ERROR	22.742	32	0.711		

WITHIN SUBJECTS EFFECTS**TEST FOR EFFECTS CALLED : TIME**

SOURCE	SS	DF	MS	F	P
HYPOTHESIS	0.675	1	0.675	0.790	0.381
ERROR	27.339	32	0.854		

TEST FOR EFFECT CALLED: Teaching Method X Time

SOURCE	SS	DF	MS	F	P
HYPOTHESIS	0.190	1	0.190	0.222	0.640
ERROR	27.339	32	0.854		

Table 8.1**OVERALL MEANS DEPENDENT VARIABLE:**

Secondary Trait Scale

	PRPT	POPTS	DIFFERENCE
CONTROL N=16	2.8	2.7	-.1
TREATMENT N=19	2.6	2.3	-.3
DIFFERENCE	-.2	-.4	

Summary Repeated Measures ANOVA: Primary and Secondary Trait Scales

The results of analysis of variance for both Trait Scales reveals a significant effect for between group differences. This effect indicates that these groups were differentially effected by the teaching method. While we expected the effect to favor the Mastery Learning group, the reverse was true. On the Primary Trait, the performance of the control group increased and that of the Mastery group remained unchanged.

On the Secondary Trait, while both the control and the Mastery groups' performance decreased, the control decreased less sharply.

Descriptive Results : Perception of Achievement, Attitude and Interest Questionnaire Phase I

The purpose of measuring affect by the use of the questionnaire was to answer one of the questions posed by the study: Do attitude and interest towards learning develop as a function of the achievement students attain and of their perception of the achievement? It was expected that students learning with a Mastery Learning approach would develop higher levels of affect than students learning with a conventional approach.

As mentioned in the previous chapter, the questionnaire was composed of three parts: Part A--Perception of Achievement--Part B, Interest Toward Learning to Write and Part C--Attitude Toward Learning to Write. The focus of the items, therefore, was aligned with the performance objectives from The Table of Specifications in both phases of the study. This alignment was done by insuring that there were representative items in each section that dealt with a) learning to write, b) grades received for writing assignments, and c) activities involved in the classroom. The same version of the questionnaire was administered three times at the completion of each instructional unit and in all classrooms.

Table 8.2 presents the averaged means and standard deviations of each section and time administration of each questionnaire. As can be noted, the overall means for Part A are almost the same for the first two time administrations for both the experimental and control, but the means for the treatment group decrease sharply relative to control's by the third questionnaire administration. Contrary to our expectation, by Time 3, the control group appears to have a relatively better Perception of Achievement than expected. The standard deviations for both groups in

Part A are roughly equivalent. For Part B, the treatment group starts out with better attitudes than the control and even improves by the third time administration. Moreover, the treatment group has lower standard deviations. For Part C, (interest) the treatment also had a better lead relative to the control and extends that lead by the third time administration, but also has higher standard deviations. Notice that the treatment group's standard deviation also fluctuates more than the control's.

Table 8.2 Phase I

MEANS OF REPEATED MEASURES:

Part A: Perception of Achievement.

Part B: Attitude and Part C: Interest Questionnaire

	Time 1 Mean	Time 2 Mean	Time 3 Mean
Part A			
Control	1.9 (.43)	2.0 (.43)	2.0 (.46)
Treatment	1.9 (.47)	2.0 (.51)	1.8 (.47)
Part B			
Control	1.5 (.89)	1.5 (.84)	1.6 (.78)
Treatment	1.8 (.26)	1.7 (.63)	2.0 (.48)
Part C			
Control	1.9 (.44)	1.9 (.50)	1.9 (.46)
Treatment	2.2 (.48)	2.1 (.63)	2.3 (.55)

*Note standard deviation in parentheses.

Summary Phase I: Perception of Achievement,
Attitude and Interest Questionnaire

The Perception of Achievement seems to be maintained constant by the control group. The treatment group's Perception of Achievement declines over time, however. The Attitude Toward Learning to Write and Interest in Learning to Write is more positive for the treatment group over time than for the control group. Although a slight trend in direction is observed, any interpretations of these descriptive results might be spurious though. As mentioned in the validity section, the analysis of the questionnaire results was laden with problems, thus patterns of responses were not easily discernable.

Formative and Summative Tests Phase I

So far we have seen that the expectations we had for the mastery group's achievement were not met. The second expectation was that the Mastery group would have higher levels of affect than the control. Generally this expectation was not met either. Still it is difficult to tell why, so let's turn then to the Formative and Summative Test results.

The objective of the Feedback and Corrective Loop is to assure that students enter each subsequent learning task with necessary cognitive entry behaviors. Since the control and treatment groups in Phase I were comparable groups (both from the same school), a comparison of scores on each item of each Formative test was made to determine achievement differences between these two classrooms. These item scores were then converted into overall means for each test. As we have seen, this loop did not appear to operate as judged by observation. Recall that we pointed out that by the second week of implementation of the study, ongoing observations in the treatment classroom revealed that Formative Tests and correctives were not being used to diagnose levels of learning in order to modify instruction or meet individual learning needs. The instruction was constantly geared down since the expectations the teacher and instructional aide had of the students were very low to begin with. Students were never grouped according to how they had performed on the first Formative Test; instruction was done on a group basis and the same level of correctives were given to the entire group not just individual groups of students. No attempt was made to use the results of the first Formative Test to provide individual feedback. The consequence is that overall achievement is not higher in the Mastery treatment group than in the conventional.

Table 8.3 summarizes both Formative and the Summative Tests used in Phase I. As can be seen, at least in terms of means, the treatment group started worse to begin with. The treatment group

closes the gap on control by the end of second Formative Test, but doesn't close the gap as much as one would expect, nor obviously as much as necessary. By the time the Summative Test was administered the gaps were very large.

Table 8.3

Summary of Phase I Formative and Summative Tests

	Control			Treatment		
	Mean	sd	Total Items	Mean	sd	Total Items
Formative Test I (Parts A & B)	.72	.24	17	.54	.26	16
Formative Test II (Parts A B & C)	.91	.08	17	.79	.07	14
Summative Test	.92	.92	4	.64	.23	4

As illustrated the treatment group in this study performed consistently lower than the control group on both Formative Tests and on the Summative Test. The standard deviation for the treatment group was slightly higher only for the first Formative test. The second Formative test as well as the Summative Test show lower standard deviations for the treatment group than for the control group. The pattern of mean achievement illustrated by Table 8.3 then, provides support for the proposition posed in the methods Chapter procedures section of this document, that the Feedback/Corrective Loop that has been so central in explaining the success of Mastery Learning versus conventional instruction broke down in Phase I.

Summary Phase I: Findings

An overall means impression appears to show that at entry skill level both groups are approximately equal. However, the repeated measures analyses of variance for the CSE Subscales show that groups were differentially affected by the treatment across time. Although groups were roughly equivalent at Pretest skill, by Posttest the control group increased while the treatment group decreased on all CSE

Measures. The analysis for the Trait Scales showed significant effects between groups; thus, the groups were affected by the teaching method. On the Primary Trait the control increased and the Mastery group remained unchanged, while on the Secondary Trait both groups decreased in performance. These achievement results were also supported by the two Formative and the Summative Test. The treatment group performed consistently lower on all tests. The affective measures were difficult to interpret, but an examination of the means for each time administration of the questionnaire revealed a slight shift in means in favor of the control group for Perception of Achievement. For Attitude and Interest, however, the treatment group means were higher than the control's across time.

Phase II

Basic Data

What we have seen so far is that the Feedback/Corrective Loop in the treatment classroom for Phase I did not really function. Perhaps if the treatment had been properly implemented the achievement results for this group may have been different. In Phase II subjects were measured under only one condition: Mastery Learning. However, in this Phase effort was made to ensure that Mastery Learning was well implemented and especially the Feedback/Corrective Loop. Matched pair t-test analysis was conducted for each dependent variable used to gauge achievement in writing skills. Table 8.4 presents the results of the CSE Subscale measures and the Trait scale measures of the Pre and Posttest scores for Phase II of the study.

Table 8.4

Phase II Means and Standard Deviations: CSE and Trait Scales

	<u>Pretest</u>			<u>Posttest</u>		
	<u>n</u>	<u>Mean</u>	<u>sd</u>	<u>n</u>	<u>Mean</u>	<u>sd</u>
CSE 1: General Impressions	30	4.0	1.2	30	4.2	.90
CSE 2: General Competence	30	3.2	1.2	30	3.5	1.2
CSE 4: Paragraph Coherence	30	3.0	1.2	30	3.4	1.4
CSE 6: Mechanics	30	2.7	.89	30	3.4	1.4
Primary Trait : Main Idea	30	3.6	1.1	30	4.3	.65
Secondary Trait: Inventive	30	3.7	.95	30	3.9	.71

Whereas in Phase I there was typically little change in means for the control group and the treatment groups' means regressed from Pre to Posttest, in Phase II there appears to be considerable change since the means for all scales increased. The overall impression is that means for Posttest scores on all measures are consistently higher than the Pretest means. Moreover, whereas in Phase I the magnitude of standard deviations fluctuated among the groups and various scales, in Phase II the standard deviations are lower at Posttest level for three scales (CSE subscale 1, Primary and Secondary Trait). The CSE 2 Subscale's variance remained the same. The CSE 4 Subscale's variance increased slightly and the CSE 6 Subscale's increased most.

Data Analysis Phase II

Table 8.5 shows that the matched t-tests in phase II reveal significant results ($P < .01$) for three of the dependent variables; CSE 4, CSE 6, and Primary Trait. Indeed given the conditions for the

Table 8.5 Phase II

Matched Case t-Tests

	<u>t</u>	<u>P</u>
CSE 1: General Impressions	.499	NS
CSE 2: General Competence	1.368	NS
CSE 4: Paragraph Coherence	2.925	.008*
CSE 6: Mechanics	3.739	.001*
Primary Trait Scale: Main Idea	3.367	.003*
Secondary Trait Scale: Inventive Language	.738	NS

implementation of Phase II of the study it is not surprising to see that these three important scales reveal significant findings. Significant difference from Pre to Posttest for the analytical CSE Subscales (CSE 4 and CSE 6), are interesting to note in light of the fact that the conditions for Phase II were conducive for the proper implementation of a bilingual (Spanish/English) Mastery Learning treatment. In addition, and even more important, is the significant result for the Primary Trait Scale, since the Primary Trait criteria was developed based on the specificity of the performance objectives stated in the Table of Specifications. In the study, the Primary Trait presumes to gauge level of acquisition of the basic unit of thought in writing; the development of a main idea paragraph. These findings show some promising trends for the implementation of Mastery Learning for teaching basic writing skills in a bilingual situation. The results of the all scales increased over time, but of important significance is that taken together all scales promoted a higher level of paragraph writing skill acquisition.

Descriptive Results Phase II: Perception of Achievement, Attitude, and Interest Questionnaire

As previously mentioned the questionnaire administered in Phase II was the same version of the questionnaire administered in Phase I. The same problems encountered with the analyses and the interpretation in Phase I hold true for Phase II (refer to the previous findings section for Phase I and the reliability section of this Chapter for a more comprehensive explanation). Nevertheless, Table 8.6 shows the means and standard deviations of the three separate administrations of this questionnaire to the single Mastery Learning group in Phase II.

Table 8.6

Phase II MEANS OF REPEATED MEASURES
 Part A: Perception of Achievement,
 Part B: Attitude and Part C: Interest Questionnaire

	Time 1 Mean	Time 2 Mean	Time 3 Mean
Part A Perception of Ach.	1.9 (.45)	1.9 (.44)	2.0 (.47)
Part B Attitude	2.0 (.37)	1.9 (.36)	1.9 (.41)
Part C Interest	2.0 (.56)	2.0 (.45)	2.2 (.40)

*Note standard deviations in parentheses.

Table 8.6 shows that the means for the three administrations of the Part A--Perception of Achievement increase slightly from Time 1 to Time 3. The standard deviations for Part A decrease for Time 2 but increase by Time 3. Recall that for Phase I, the means for Perception of Achievement for the control group went up over time, but that for the treatment group they went down. This pattern was also true for the standard deviations; the control group went up, while the treatment group's went down by Time 3 in Phase I.

The Part B--Attitude by the third time administration shows a slight but gradual decrease in means, and there is an increase in standard deviation. This is quite a different pattern than for Phase I. Recall that in Phase I, the Part B--Attitude means for the control group went down while for the treatment group went up by the third time administration. Also recall that the standard deviations for Attitude (Part B) for Phase I went down for both groups by Time 3.

The Phase II means for Part C--Interest increase with each successive administration, and the variance decreases considerably (remember that just as in Phase I the Part C section of the questionnaire had the highest reliability). Again recall that the means for Phase I Part C for the control group remained the same and the treatment group means increased. While both groups in Phase I had lower standard deviations by Time 3, they were not as low as the standard deviations for Phase II Part C.

Summary Phase II: Perception of Achievement, Attitude and Interest Questionnaire

In summary, Phase II results of the Perception, Attitude and Interest Questionnaire of this study appear to show that there was a slight increase for Perception of Achievement, Attitude did not change much and, in fact, decreased slightly but Interest in Learning to Write increased over time.

Summative Tests for Phase II

So far we have seen a different pattern for Phase II of the study than for Phase I. In Phase I we traced why the treatment group had performed lower than the control group by Posttest level. The results confirmed the proposition that the Formative Test process broke down in the treatment classroom and thus resulted in a lower achievement level for the Mastery Learning group. Let's do the same for Phase II. Since the overall pattern is different we expect the results of the Summative Tests to also be different. As with the Formative and Summative results of Phase I of the study, a comparison of scores on each item of each Summative Test was made to determine achievement differences for Phase II. Just as in Phase I, the objective of the Feedback and Corrective Loop in Phase II is to assure that students enter each successive learning task with prerequisite cognitive entry behaviors. We have seen in the mean results of the achievement measures that this process has occurred. Now let's examine the results of three unit cumulative skill Summative Tests. Table 8.8 shows the means and standard deviations for each of the instructional units in Phase II.

Table 8.7

Summary Phase II:
SUMMATIVE TESTS RESULTS

	Mean	sd
Summative I	.86	.10
Summative II	.88	.16
Summative III	.98	.02

As can be seen from Table 8.7 the means and standard deviations for Summative Test III reveal that achievement increased and generally variability in performance decreased with each successive administration of a cumulative Summative Test. Unit by unit achievement measures reveal that in Phase II subjects' performance was high on all measures. Overall achievement was above the set 85% criterion level for Mastery when scores for all items were combined. By the time the third test was administered the means were quite high, and there was very low variation. This is evidence that all students in Phase II of the study were performing and learning at a higher level of skill. Standard deviations were generally low for all measures.

Summary Phase II: Findings

The overall first impression from an examination of the means for Phase II is that performance was high on all measures from Pre to Posttest. There was also less fluctuation on the standard deviations than in Phase I. The two analytical subscales (CSE4, CSE6) had higher standard deviations at Posttest level. Indeed, the examination of mean results are substantiated by the results of the t-test analysis. There were significant results for three scales: CSE 4, CSE 6, and Primary Trait.

The results of the Affect Questionnaire for Phase II revealed a slightly different pattern than those of Phase I. Perception of Achievement was higher over time, Attitude remained constant, but Interest in Learning increased. The standard deviations for both

Perception of Achievement and Attitude increased over time. Standard deviations for Interest in Learning sharply decrease however, by the third time administration of the questionnaire.

The final evidence for the success of Mastery Learning in Phase II and, hence, overall increase in achievement are the results of the unit by unit Summative Tests. The Summative Tests results show that there was a continual increase in achievement from unit to unit and that overall variance sharply decreases over time.

CHAPTER VII

Discussion

As noted at the onset of this dissertation the problems of illiteracy and dropping out of school among Hispanics in this nation are deplorable. The phenomenon is complex and difficult to understand given the factors attributed to leaving school early for Hispanic youth and particularly for Mexican American/Chicano youth. Most of what is presently known about Hispanic non-achievement shows that compensatory programs designed to meet the needs of this population have not been successful. The high dropout rate of Hispanics in California is indicative of the negative attitude that students have toward further schooling. It has been concluded that low teacher expectations, irrelevant and rigid instruction, tracking, and other unjust practices towards Hispanics engender a negative school environment. The motivation to learn is low or, in most cases, non-existent due to the extremely competitive classroom atmosphere. The struggle for achievement is related to how students perceive themselves (academic self concept) and how teachers evaluate their progress in comparison to others. The literature points to several principle factors that influence learning and the motivation for learning in the domains of school: teacher attitudes and expectations which are manifested in competitive classroom environments, and testing-evaluations procedures. All factors are interrelated and have all been shown to affect Hispanic students' achievement.

This study explored one educational way of breaking the vicious economical, social, and political cycle into which Hispanic illiteracy has fallen. Hence this study has examined one approach for changing the conditions of the classroom environment in order to promote a higher level of achievement and a higher level of academic

self-esteem. This approach was examined in two naturally occurring public school classroom situations and used one grade level--the fourth. Specifically, this study examined a Mastery Learning approach in comparison to a conventional approach at one school and a Mastery Learning only condition at another school for teaching basic paragraph writing skills. The intent of this study was to examine the degree to which a "Mastery Learning" approach, as an equity structure, is a better approach for teaching and learning than a conventional approach (which stresses competition) for students of Hispanic background. The focus of a Mastery Learning approach was on the importance of sufficient rewards for all students, irrespective of ability level. The Mastery Learning approach proposed that interest and Attitude Toward Learning could be altered positively for all students regardless of placement or prior achievement.

In order to test the applicability of a Mastery Learning equity structure for teaching basic paragraph writing skills to students of equivalent academic ability, but different levels of English language proficiency, recall that the study utilized the theoretical model developed and validated by Anania (1983) which compared three learning approaches (see Figure 3.1, Chapter III). The present study utilized three basic components of Anania's model: Antecedent Conditions, Instructional Approach, and Learning Outcomes (see Figure 3.2 Chapter III for Phase I and Figure 4.1 Chapter IV for Phase II).

Bloom's Taxonomy was used to delineate the writing instructional tasks and writing objectives that the students of the study would be taught and also to identify appropriate instructional material to be used. Next, several measures of writing achievement and one measure of affect were developed. These were the modified versions of four Center for the Study of Evaluation (CSE) Subscales, the development of two Trait Scales which were used to gauge

paragraph writing achievement at the Pre and Posttest level, and a Perception of Achievement, Attitude and Interest Questionnaire. In addition, Formative and Summative Tests to be used as diagnostic process measures, were developed as well as additional instructional material used for corrective/feedback activities. In order to gauge achievement, a Pretest Writing Sample was gathered to determine entry skill level. A Perception of Achievement, Attitude and Interest Questionnaire was administered shortly thereafter. Then the instructional sequence began culminating after a four week period of time. During these four weeks, an ongoing diagnostic process was followed by the use of Formative and Summative Tests and two additional administrations of the same Perception of Achievement, Attitude and Interest Questionnaire. As a final measure of achievement, a Posttest Writing Sample, was gathered.

Essentially the study had two designs since it was conducted in two phases. Phase I of the study (see Figure 3.3, Chapter III) compared two instructional conditions: a Mastery Learning approach and a conventional instruction approach. The study's Phase II design only had a Mastery Learning approach condition with no comparison group (see Figure 4.2, Chapter IV) and was administered at a different school site. Phase I of the study was initiated in September and finalized in November of 1986 and Phase II of the study was initiated at the end of November 1986 and finalized on January 1987. Writing samples for Pre and Posttest, Formative and Summative Tests and affect questionnaires were gathered from a total of 84 students for both phases of the study

Summary of Findings

The analysis of data for this study was done in several stages: first the instruments for gauging the Pre and Posttest samples were selected modified and pilot tested. Next, interrater reliabilities were

done on the scores obtained from the ratings conducted by using the four CSE Subscales and the two Trait Scales. Correlations for all interrater reliabilities were high (.90 or higher) with the exception of two Pretest Reliabilities which were at .85 and .86 for the two most analytical Scales (Subscales 4 and 6). Then the consistency of the internal reliabilities for all instruments used was checked by using Cronbach's Coefficient Alpha as the measure for score reliability. The Pre and Post Achievement Measures tended to have the greater reliability than the affect ones. The different sections of each Formative and Summative Tests in Phase I varied in length and content; thus it can be surmised that the reliabilities of these measures are questionable. Similarly, the reliability of the Summative Tests in Phase II (recall that only Summative scores were used for final gauging of achievement in this phase) was rather low (.35). The internal reliabilities of the affect questionnaire also had questionable reliabilities.

The next procedure followed in the analysis was to examine the construct validity of all instruments used in the study. The procedures used were iterative. Pearson Product-Moment Correlation Coefficients and related Principal Components analysis were used to compare the pattern of correlations as well as the magnitude of the correlations among variables and factors in the data sets. The Pre and Posttest Achievement measures had adequate construct validity. For both the CSE Subscales and Trait Scales, there appeared to be high moderate correlations among the Subscales and low moderate correlations between the Subscales and the Trait Scales. This resulted in both convergent and discriminant validity for these Pre and Posttest Measures. The Formative and Summative Tests for both phases had questionable construct validity, though, as did the construct validity of the affective measure, the Perception of Achievement, Attitude, and Interest Questionnaire. The results were affected by small sample sizes, missing data points, and

insufficient variability in responses which then prevented adequate interpretation of underlying constructs.

The next stage of analysis of the data examined the empirical relationships for each separate phase. We turn first to examine the Phase I data. Overall means and standard deviations of the Pre and Posttest Measures for Phase I were examined. The overall impression was that at entry skill level both groups in Phase I were approximately equal, but by Posttest level, the control group performed higher than the treatment on most measures. Variance was approximate for both groups but fluctuated more by Posttest level (see Table 7.0 Chapter VI).

Then repeated measures analyses of variance for the CSE Subscales were conducted. These analyses indicated that groups were differentially affected by the treatment across time. Although both groups were roughly equivalent at Pretest skill, by Posttest the control group increased while the treatment group decreased on all CSE Measures. On the Primary Trait, the control increased and the Mastery group remained unchanged, while on the Secondary Trait both groups decreased in performance. The Perception of Achievement seems to be maintained constant by the control group. The treatment group's Perception of Achievement declined over time, however. The Attitude Toward Learning to Write and Interest in Learning to Write seemed to be more positive for the treatment group over time than for the control group. The treatment group performed consistently lower on the two Formative and the Summative Tests.

The final stage of analysis involved an examination of the observed empirical relationships for Phase II of the study. The overall first impression of the means for Phase II was that performance was high on all measures from Pre to Posttest for the single Mastery group. There was also less fluctuation on the standard deviations than in Phase I, though the two analytical Subscales (CSE4, CSE6) had higher standard deviations at Posttest level. These findings were then

substantiated by the results of a series of t-test analyses. There were significant results for three Scales: CSE 4, CSE 6, and Primary Trait. The Summative Tests results show that there was a continual increase in achievement from unit to unit and that overall variance sharply decreased over time.

The affect data for Phase II were different than those of Phase I. Perception of Achievement was higher over time and Interest in Learning increased. Only attitude remained constant. The standard deviations for both Perception of Achievement and Attitude increased over time. Standard deviations for Interest in Learning decreased by the third time administration of the questionnaire, however.

The overall summary for both Phases of this study of observed empirical relationships are illustrated by Table 8.8.

Table 8.8
Summary of Observed Empirical Relationships
Achievement Measures

<u>Phase I</u>	<u>Phase II</u>
<u>CSE Subscales</u>	<u>CSE Subscales</u>
*control increased	*ML increased
*ML decreased	on all Subscales
<u>Primary Trait Scale</u>	<u>Primary Trait Scale</u>
*control increased	*ML increased
*ML remained constant	
<u>Secondary Trait Scale</u>	<u>Secondary Trait Scale</u>
*control decreased	*ML increased
*ML decreased	
<u>Formative/Summative Tests</u>	<u>Summative Tests</u>
*control increased	*ML increased
*ML decreased	from unit to unit
<u>Affect Measures</u>	
<u>Phase I</u>	<u>Phase II</u>
<u>Perception of Achievement</u>	<u>Perception of Achievement</u>
*control remained constant	*ML increased
*ML decreased	
<u>Attitude</u>	<u>Attitude</u>
*control decreased	*ML remained constant
*ML increased	
<u>Interest</u>	<u>Interest</u>
*control decreased	*ML increased
*ML increased	

Limitations of the Study

Both designs used for this study were essentially exploratory. To reiterate, the study was done in two phases and conducted at two separate schools in the same community. The conditions for both phases were naturally occurring situation, environments. The Pre and Posttests used as achievement measurements were made on a number of dependent variables.

The original design for Phase I was developed in order to account for the eight internal validity threats identified by Campbell and Stanley (1963). Design 10, a self-selected nonequivalent control group design, was originally identified for use in this study. In Phase I (as stated in Chapter III: Methodology) both classrooms were treated somewhat similarly. So it was assumed that main effects of history, maturation, testing, and instrumentation would be better controlled. Each of the threats to internal validity plus others identified by Campbell and Stanley (1963) will be discussed in turn to determine the degree of internal validity for Phase I of this study.

A One-Group Pretest Posttest Design was utilized for Phase II of this study. It is described by Campbell and Stanley (1963) as Pre-Experimental Design 2. With this design there was no control group, therefore no comparisons can be made. So, since there is no comparison group there can be extraneous variables that can jeopardize internal validity. As with the first phase of this study careful analysis of the factors jeopardizing the validity of the experimental design must be considered before drawing any conclusions from the observed results.

Internal Validity

Internal Validity of Phase I

According to Stanley and Campbell (1963) quasi-experimental designs like Design 10 (Phase I) of this study can have greater control over external than internal threats to validity. For example, use of intact groups will probably decrease subject reactivity. A self-selected design is much weaker because the groups are likely to be highly dissimilar. Selection is always a problem for this design because a researcher can never be certain that the groups are equivalent unless the subjects are randomly assigned to comparison groups. If selection is a threat, then history and maturation are also threats.

Regression and mortality can also be threats to the nonequivalent control group design. Statistical regression will be a threat to the validity of this design if either of the comparison groups has been selected on the basis of extreme scores. Thus, a difference in scores from the Pretest to the Posttest between the two groups selected for Phase I may be the result of statistical regression rather than the effect of the treatment. Design 10, however, may control for instrumentation and testing and also can control for subject reactivity (external validity). Nonetheless, let us examine these dimensions closely for Phase I.

A basic threat to the internal validity of a study is History, the change producing events occurring between the Pre and Posttest in addition to the experimenter treatment that affect the results. During the four weeks of the instructional sequence for Phase I of the study between the Pre and Posttest, the design events of History did effect the comparison groups differently. The fact that the groups were tracked into ability sections perhaps influenced the instructional activities that students were exposed to and it is likely that this promoted unwanted learning not accounted for by the

treatment. Since the low-ability section regressed, the affects of History are a questionable threat to this phase of the study.

Another threat to internal validity is Maturation if ongoing maturational processes or experience of subjects other than the "treatment" variables could account for the observed results. These processes would have to affect the comparison groups differently to be a threat to internal validity. This threat is assumed not to be a serious threat to Phase I of this study. However, since time remained constant between the control and treatment group during a four to five week period and ongoing activities were basically the same, it is very unlikely that maturational factors influenced the results.

If the process or outcomes of taking a Pretest seem to have an effect when the test was administered, Testing is then a threat to internal validity. Although this design controls for testing, it can be assumed that testing by using an attitude questionnaire initially (at the onset of the instructional sequence) was a threat in Phase I. The Affect Questionnaire was administered simultaneously to the achievement measures. It appears that this affect questionnaire, which is the type of instrument that children are not accustomed to answering, may have affected the internal validity of the study.

Instrumentation refers to autonomous changes in the measuring instruments which might account for a Pre to Posttest difference. The criterion-referenced instrumentation was clear and designated by whether the students had 'mastered' or 'not mastered' the objective of each learning unit. The gauging of Pre and Posttests was carefully monitored to insure anonymity, rater fatigue objective judgment of essay quality, and reliable score assignment (see Chapter III). Pre and Posttest Writing Samples were coded and shuffled around in order to present them to two independent raters in a non-sequential fashion. The techniques were used as control for objective gauging of the writing samples.

In spite of this meticulous effort to control instrumentation, this dimension is a threat to internal validity in Phase I of this study since changes occurred in the measurement process that were a continual practice throughout a four to five week sequence of instruction. For instance, students' performance on the unit to unit Tests may have been affected by instructional activities in other classes since the instruction for the study only took place during a thirty minute period of time during the school day.

As stated before, the study used criterion-referenced measurements to measure the main concepts of each unit to each unit Formative Test thus the items and gauging for each instrument administered were considerably different for each test administered throughout each of the five week sequences for the separate phases. The Affect Questionnaire included questions about self, other class members, teacher, and material content. Thus, while the design chosen for this phase of the study controls for testing and instrumentation whether the Affect Measures are a threat to internal validity remains an open question.

Statistical Regression is a threat to internal validity if the groups being studied have been selected on the basis of extreme scores. Regression does seem to be a threat to the internal validity of this study.

A major concern in Phase I was the way the students had been tracked into ability sections (even though the scores on the Pre test reveal that at Pretest level both groups were comparable). The Pretest means for both groups were approximately equal: there was very little discrepancy in scores between the two groups on all the dependent measures. Hence, the source of difference shown by the results for Phase I was not ability (at least as it appears according to external measures). It is likely that there were other sources of difference between the two groups, such as awareness of being tracked into ability sections, differential attitudes towards

predominate Spanish speakers, watered down lesson presentations, etc. These factors seem to influence the internal validity of this phase since this is a self-selected non-equivalent control group design, and there was no possibility for random assignment to treatment.

Selection is a concern for this study. In Phase I, it threatens internal validity since differential processes or criteria were used in selecting subjects for comparison groups. The site for data gathering was a commonly occurring educational condition in which students in the particular groups were tracked into ability sections. Random assignment of subjects was not possible for this study. As stated in Chapter III, ability placement was based on CTBS scores from previous years, previous teachers recommendations, and teacher subjective evaluation of basic language skills.

Another very important dimension regarding Selection is the varying abilities of the subjects in overall English language proficiency. Varying levels of linguistic proficiency in English is important because all of the students that were considered limited English speakers were 'tracked' into the low-ability section. The only native English speakers in this low ability section were subjects that had low-ability language skills. It should be noted, though, that the dependent measures were gauging writing skills by six dimensions. Students were asked to write paragraphs for the Pre and Posttests in whatever language they chose (either Spanish or English), they were not gauging English language proficiency, per se, but rather basic writing skills.

Yet another Selection factor that might be a threat to the internal validity is that in Phase I of the study the teachers for these comparison groups were assigned based on language skills. The bilingual teacher was assigned to the low-ability section even though the objective was not to use both languages for instruction, but to transition all students to an English only instructional curriculum. The random assignment of this factor also was not feasible for this

phase of the study. The selection dimension of internal validity does seem to pose a threat to Phase I of this study.

Mortality (biological, psychological, or systematic factors that vary with the passage of time) has been identified as a threat to internal validity. The conditions of the study were such that Mortality is not a threat for Phase I. There were no data that showed any major discrepancies of a dropout problem in the comparison groups thought there were still many daily student absences. The high absenteeism rate is the reason for missing data points throughout the study on different measurements. But all students with the possible exception of one, completed some aspects of the five week instructional sequence in this phase.

Internal Validity of Phase II

In Phase II (Design 2) of the study a four week instructional treatment elapsed between the Pretest and the Posttest, and although there was no clear evidence of changes, events (History) may have occurred in addition to the treatment that may have caused a difference. For instance, the fact that Phase II was implemented around Christmas time may have influenced the results since children are usually more optimistic and positive at this time of the year. Thus, History was a threat to internal validity of this phase of the study.

In addition Maturation could have been affected by the season or institutional events that were scheduled at this time of the school year. Although nothing unusual was observed, such events might produce changes between the Pre and Posttest that are confusable with the effects of the treatment. Maturation is seen as an unlikely threat to the internal validity of Phase II however.

Also in Phase II, as with the previous phase of the study, the testing done was unobtrusive and was very much related to the

instructional material. Testing though is seen as a threat since the Pre and Posttest were exactly the same in content. When tests of this sort are used, in general there is better performance on the second test. In addition, the affect questionnaire is seen as an intrusive measure and may have affected the students performance on the Pre and Post Measures.

The design for Phase II does not include a comparison group. The group used was not selected based on extreme scores. It was selected because the teacher volunteered. So Regression to the mean does not seem to pose a threat to the internal validity of Phase II.

In Phase II the teacher volunteered to implement the treatment in her already assembled bilingual classroom. The placement of students was done objectively based on previous assessment of language skills. The criteria used to determine suitability for this study was that the subjects in this group were fourth-grade students and that the linguistic composition was diverse (from monolingual English to monolingual Spanish and various levels of bilinguality along this spectrum) and that instruction was bilingual. Selection though, does pose a problem for the internal validity of Phase II.

Mortality is not considered a threat to internal validity either; at least with respect to overall participation in Phase II.

External Validity

Campbell and Stanley (1963) identify four potential threats to external validity. External validity attempts to answer the question as to what population, treatment variables and measurement variables can the observed effect be generalized. Bracht and Glass (1968) expand on Campbell and Stanley's descriptions and identify ten potential threats, then divide them into two areas; population

validity and ecological validity. First the population validity will be dealt with.

External Validity of Phase I (Population)

Interaction of Testing and Treatment is only a threat to external validity if such interaction has a semblance to the hypothesized effect of the 'treatment.' This is a threat in Phase I since the treatment was an on going instructional sequence. Although there were different forms of Tests available to use, the treatment was not properly implemented in Phase I because not all of the testing material was utilized.

Another possible threat to the external validity of this study is the Interaction of Selection and Treatment. This interaction is a strong possibility in Phase I since the subjects had been tracked into ability sections. The treatment was implemented in the 'low ability' section. The sloppy classroom implementation of the treatment in Phase I coupled with the fact that the two groups were not equivalent cohorts sheds doubt to the external validity of this study along this interaction dimension.

There is some possibility that some 'arrangements' of the study were reactive in Phase I. Reactive Arrangements as threats to external validity are arrangements, context, or procedures of the study which are obviously (to the subjects) unrepresentative of the contexts (such as classroom or course settings) to which the investigator wishes to generalize the results. It is quite possible that the Affect Questionnaire was reactive since most of the items dealt with how the students reacted to self, teacher, and other students in the classroom context.

Another dimension of how the 'arrangement' caused reactive responses in Phase I was the treatment teacher's attitude towards the procedures and responsibilities after the initiation of the project.

Furthermore, the majority of the subjects that comprised the treatment groups were part of a program designed to assist them in making the transition from native language Spanish skills to the use of English skills as a formal language of instruction, hence, there were many levels of English language proficiency skills in one group. This circumstance was in part the reason the treatment teacher might have felt overwhelmed and why she developed a negative attitude. This situation suggests that Phase I results can only generalize to members of this special population rather than to some larger population, for instance, to all fourth grade students in California. The reality however, is that this special population is a common group, especially in places where there is a large, bilingual Spanish language community. This study was designed as exploratory for purposes of testing a Mastery Learning approach for teaching basic writing skills to Spanish language background students. Careful specification of what type of Spanish language students was not a primary concern given the diversity of this population with such factors as SES level, immigration status, English language proficiency level, basic skills levels in either language, and cultural background.

One final but important dimension of the population external validity of Phase I of this study is Interaction of Personological Variables and 'Treatment Effects'. Interactions between the treatment variable and characteristics of the subjects may limit the generality of the inference depending on the type of interaction. Linguistic proficiency in English maybe a threat to Phase I. However, the results of the Pretest indicate that both groups in Phase I were comparable in writing skills regardless of language used for writing (Spanish or English).

External Validity of Phase II (Population)

Design 2 utilized a Pretest and in fact the Affect Questionnaire was administered directly after the Pretest; thus, the Interaction of Testing and Treatment is a threat to the external validity of Phase II. It is quite likely that the students' perceptions were affected by the questions posed on the affective questionnaire. In addition, although testing is a regular activity in classrooms, fourth grade students are not accustomed to responding to questions regarding their attitudes.

Since the Phase II design only included one single group at another school site than Phase I, the contextual conditions were quite different. Although the schools used for both Phase I and Phase II were in the same community, they were distinctly different. For example, the classroom site used for Phase II offered bilingual instruction for enrichment purposes on a volunteer basis. This was not the case for Phase I; they offered transition classes for non-English speakers. Thus the Interaction of Selection and Treatment is questionable on the basis that the characteristics of the school may have caused the experimental treatment to be more effective (the Phase II teacher enthusiastically volunteered).

In Phase II, it is quite possible that the Affect Questionnaire produced a Reactive Arrangement since most of the items dealt with how the students reacted to self, teacher, and other students in the classroom context. This fact may be a threat to the external validity of Phase II. Another dimension is that the experimenter participated in the actual instruction. Although the researcher's participation was supposed to be unobtrusive, whether or not her presence created unobserved reactions on the part of the students remains an open question.

Just as in Phase I, in Phase II linguistic proficiency in this study is considered a personological variable. There were varying

levels of proficiency Spanish and English in the same classroom. Thus, the Interaction of Personological Variable and Treatment is a threat to external validity for Phase II. The treatment group in Phase II had high Pretest scores that increased on all measures by Posttest. This result may be due to the treatment, but also to several other factors, such as, bilingual instruction, the time of initiation of the treatment (December thru January), and the teachers attitude toward bilingual students.

The ten threats that Bracht and Glass consider in the Ecological Validity dimension are the following:

Ecological Validity Phase I

Describing the Independent Variable Explicitly is necessary for replication of experimental results. This is not a threat to Phase I since there was considerable care taken at specifying all aspects of the treatment and experimental setting.

Multiple-treatment Interference occurs when multiple treatments are administered to the same subjects. One treatment that was sequential and continuous and which was measured by criterion-referenced instruments was administered in Phase I. Even though Mastery Learning is seen as one treatment, the fact that measurement was continuous could appropriately be seen as multiple treatments. The instruction in the control classroom could also be seen as a "treatment" of sorts. Therefore, this dimension could be considered a threat to the external validity of this study.

The Hawthorne Effect occurs if the behavior of the subjects is altered by their perceptions of being in an experiment. When this happens it is a threat to ecological validity. The subjects in this study were aware that they were involved in the study. The material content of the study however appeared as typical instructional material since the resources available in the classroom were used to

develop the treatment material. It seems unlikely from the observed relationships that this would affect the subjects. The administration of the affect questionnaire, three consecutive times during the study is questionable. Since fourth-grade students typically are not often asked to respond to these types of questionnaires, there is a basis for claiming that the task made students aware that they were part of an ongoing experiment. This dimension remains an open question that could produce perceptions or behaviors that become a threat to ecological validity of this study.

Novelty and Disruption Effects is a dimension that affects ecological validity if enthusiasm or disruption generated by the newness of the treatment affects the results. Again, as with the previous threat, the content of the treatment was not novel in that material for the treatment was generated from the available sources in the classroom, but it is unknown as to whether the disruption of the administration of the affect questionnaire influenced the results in Phase I.

Experimenter Effect is a threat when characteristics or behavior of the experimenter influences subject behavior. In Phase I there was a conscientious effort by the experimenter to become inconspicuous when observing classroom procedures. Initially, in the treatment classroom, the experimenter sat in the teachers planning space developing materials while intermittently observing. The ongoing presence of the experimenter (at least four days out of a five day instructional week) also reduced this effect. The observations in the control class in Phase I were not so frequent since the teacher preferred not to be observed on an ongoing basis. Even though all tests and textbook materials used were the same, there was an opportunity to observe this control group only three times during the course of the four to five weeks.

Another factor which is considered for both phases of this study is that the subjects were quite accustomed to different adults

'coming and going' in this type of instructional setting since there were aides, tutors and other auxiliary personnel constantly present. Also, the fact that the groups of students change for different ability groups throughout the day (i.e., spelling, ESL, reading groups) further reduced the effects of this type of threat to the external validity of this study.

Pretest Sensitization is a very likely threat to the external validity of this study. In order for the Pretest to be a threat, the results would have to partly reflect a sensitizing effect. The writing Pretest was a task very similar to typical short writing assignments used in this type of classroom setting for both phases. The initial Affect Scale however, may have created some Pretest sensitization about how students felt about themselves in the classroom, about teachers, and their attitudes and interest towards the content of the instructional treatment.

Posttest Sensitization may have occurred and renders the post experiment test as a threat to the ecological validity of this study. It seems possible that the administration of the Affect Questionnaire at the end of the final unit of instruction may have affected the results of the Achievement Posttest in both phases. It is difficult to gauge just how much sensitization the questions posed in the questionnaire affected students feelings about their own ability and which influenced their performance on the Posttest. It seems that what Bracht and Glass describe as latent or incomplete effects of post sensitization would, to some extent, affect students' feeling about their performance.

Interaction of History and Treatment Effect is a threat to the ecological validity of this study. There is a strong possibility that particular events which occurred during the study differentially affected the comparison groups. In Phase I, there were several events concerning the proper implementation of the treatment (i.e., negative attitude about the treatment, no sequential follow-up of

treatment tasks, low expectation of student performance, etc.) that, though observed and dealt with by the experimenter, definitely affected the results of this study. Therefore, it is almost certain that for Phase I of this study this interaction was very strong.

Measurement of the Dependent Variable is a questionable threat to validity. The identification and selection of instruments to measure the dependent variables in this study were carefully monitored and are not considered threats to the external validity. The instruments selected and developed to measure the dependent variables produced very high interrater reliabilities and strong convergent as well as discriminant validity. In addition, they were carefully implemented. Also, the main dependent variable, the Primary Trait Scale, was directly related to all aspect of the instructional sequence based on the Table of Instructional Specifications. There is a question however, concerning one of the Scales used. The dependent measure the Secondary Trait Scale (the use of inventive or elaborative language) was developed to assess the presence of creative language in the writing samples. The experimenter feels that it was perhaps too ambitious to measure this aspect of writing for Phase I of the study.

Interaction or Time of Measurement and Treatment Effects is a threat to external validity if a treatment effect, which is observed immediately after the treatment period, may not be maintained at some later time. This dimension is a likely source of invalidity for phase I of this study. First, the treatment is an instructional sequence (Mastery Learning) that if implemented properly has an inherent mechanism for the likely retention of skills. The progressive learning of pre-requisite skills before continuing to the next skill level and the feedback and corrective dimension provide the learner the opportunity to 'master' the skills. Unfortunately, the treatment was not properly implemented and thus becomes a threat to the external validity.

Time, however, was maintained constant in Phase I between the treatment and the control group. The measurement at the end of a four week period of instruction reflects a progressive periodic measurement of skills along a continua of difficulty. Theoretically, the assumption is that if skills are learned through 'mastery', the learner will be proficient and could perform these skills at any time after Mastery Learning takes place. Therefore, if the skills were measured at subsequent four week intervals, more than likely the retention level would be quite high. Second, the content of the treatment was writing a basic paragraph. The measurement variables were six dimensions related to different aspects of a total unit: a single paragraph. If the different elements of a paragraph are learned well, then the ultimate product is a cohesive paragraph containing, at minimum, five dimensions of the skills measured by the dependent variables. Finally, in Phase I, it is unfortunate that the improper implementation of a Mastery Learning approach prevented any meaningful analysis of the factors relating to validity of the study since general conclusions cannot be drawn from the observed results. The improper implementation of the treatment does not allow generalizations about the effects of treatment to be made since the divergent results of the analysis clearly are a problem.

Ecological Validity Phase II

Describing the Independent Variable Explicitly, just as in Phase I, was not a problem in Phase II. The independent variable, Achievement with Mastery Learning, was explicitly described and is not considered a threat to external validity.

Multiple-Treatment Interference occurs when multiple treatments are administered to the same subjects. Just as in Phase I, one treatment that was sequential, continuous, and measured by

criterion-referenced instruments was administered in Phase II. Even though Mastery Learning is seen as one treatment, the measurement was continuous and could appropriately be seen as multiple treatments. Therefore, this dimension could be considered a threat to the external validity of Phase II.

The subjects in this study were aware that they were involved in a research project; therefore, this could produce the Hawthorne Effect. The material content of the study appeared as typical instructional material because the resources available in the classroom were used to develop the treatment material. However, it seems likely from the observed relationships that the researcher's involvement in classroom instruction in Phase II would affect the subjects. Also, the administration of the Affect Questionnaire three consecutive times in the duration of the study, is a questionable threat to the external validity.

As with Phase I Novelty and Disruption Affects are a questionable source of external invalidity. Although the content of the treatment was not novel in that material for the treatment was generated from the available sources in the classroom, it is unknown as to whether the disruption of the affect questionnaire administered influenced the results in Phase II.

The effects of the Experimenters Presence is a plausible threat to external validity in Phase II since the experimenter participated in all classroom activities. Although the conditions that were true for Phase I hold true for Phase II, this threat is possible.

Pretest Sensitization and Posttest Sensitization for Phase II was the same as for Phase I.

Interaction of History and Treatment Effect is a likely threat to the ecological validity of Phase II since the treatment was administered during the Christmas season. For reasons discussed in the internal validity section for Phase I this dimension is considered a plausible threat for external validity.

For the same reasons described in Phase I, ecological external validity Measurement of the Dependent Variable was not a threat for Phase II of this study.

Representativeness of Phase I and Phase II

It is imperative to account for variation in behavior when conducting research in natural settings. To reiterate, since random assignment was not possible, already assembled groups and already assigned teachers to these groups were used for this study. Snow's (1974) model of quasi-representation in educational research provides a model for analysis, "to see whether the experiment fits the nature of the behavior being studied and whether it includes the means for discovering this fact." This model of quasi-representativeness considers three dimensions: population representativeness (students) which includes studies with minimal to maximal subject characteristics description; ecological representativeness (treatment) which includes settings from laboratory to 'natural' classroom settings; and referent generality which concerns increasing learning outcome measures. In this section each of these dimensions will be considered with respect to both phases of this study.

Quasi Representativeness Phase I and Phase II

In terms of Population Representativeness this study presents an adequate description of the characteristics of the sample used. To obtain baseline data, a Pretest was utilized as a measure and furthermore, the study also provides information about how the groups that constitute the samples were assembled. In addition, in the literature reviewed, the sample of subjects is dealt with in theory and in prior research which accounts for the reasons that the instructional situations are reasonably relevant to this experiment.

Ecological Representativeness in this study was towards the 'more representative' end of Snow's continua. Two separate designs were utilized (Phase I Non-Equivalent Control Group and Phase II Pre-Post One Group) in multiple school conditions, measurement was unobtrusive (with the exception of the affect questionnaire), and the duration of the study was adequate in both phases (four sequential weeks of instruction and approximately 12 hours of total instructional time). In addition, a very important element for quasi-representativeness with respect to the treatment characteristics was that the instructional methods required covariation. The characteristics of a Mastery Learning treatment allow for multivariation and are interrelated and continuous. These factors were observed when students are provided with feedback and correctives. A naturally occurring learning process appears to take place, especially for basic learning since pre-requisite skills are necessary for higher level skill acquisition.

The representativeness in terms of Referent Generality according to Snow's dimensions places this study toward the 'more general' end of the continua. The dependent variable measures include various aspects of what constitutes paragraph writing skill. Analytical (CSE 4, 6) as well as 'holistic' (CSE 1, 2 and Trait Scales) measures are used to assess mastery skills in writing. In addition, affective factors were gauged by the use of a questionnaire (although believed to be obtrusive) that included Perception of Achievement, Attitude and Interest Toward Learning. According to Snow's model, the more the measures used are representative of several regions of reference, the more the study fits into the 'more general' dimension of quasi-representativeness.

Table 8.9 illustrates a summary of the study's limitations for both phases according to guidelines from three sources. As can be seen by the summary, this study's ecological validity is stronger than its internal or external validity.

Table 8.9

Summary of Study Limitations for Phase I and Phase II

<u>Internal Validity Stanley and Campbell (1963)</u>	<u>Phase I</u>	<u>Phase II</u>
History	?	-
Maturation	+	+
Testing	-	-
Instrumentation	-	-
Statistical regression	-	+
Selection	-	-
Mortality	+	+
<u>External Validity Bracht and Glass Population (1968)</u>	<u>Phase I</u>	<u>Phase II</u>
Interaction of testing and treatment	-	-
Interaction of selection and treatment	-	?
Reactive arrangements	-	-
Interaction of personological variables and 'treatment effects'	?	+
<u>External Validity Bracht and Glass Ecological (1968)</u>	<u>Phase I</u>	<u>Phase II</u>
Describing the independent variable explicitly	+	+
Multiple-treatment interference	-	-
The Hawthorne effect	-	-
Novelty and disruption effects	?	?
Experimenter effect	+	-
Pretest sensitization	-	-
Posttest sensitization	-	-
Interaction of history and treatment effect	-	-
Measurement of the dependent variable	+	+
Interaction of time of measurement and treatment effects	-	-
<u>Quasi-Representativeness Snow's model (1974)</u>	<u>Phase I</u>	<u>Phase II</u>
Population representativeness	+	+
Ecological representativeness	+	+
Referent generality	+	+

*Note: + means validity is controlled, - means validity is not controlled, and ? means validity is questionable

The Validity of Mastery Learning Studies

In recent reviews, the validity of Mastery Learning programs have been tested, challenged, and analyzed in a variety of ways, especially by synthesizing the characteristics of the studies into so-called meta analyses (see Guskey and Gates 1987; Kulik, Kulik and Bangert Downs, 1990; Slavin, 1987, 1990). In order to assess whether the present study includes characteristics identified by these researchers and accounts for some of the issues

being scrutinized, an examination of salient features from two sources follows.

According to C. Kulik, J. A. Kulik and Bangert-Downs R.L., (1990) the meta analysis studies with large effect sizes have a social sciences rather than math, natural science, or humanities content. Also, they were likely to use locally developed rather than nationally standard tests as criterion measures of student achievement. The mastery level criterion was set at 100% performance on tests and students were required to do more through course material at the teacher's pace, not at individual student rates. Finally, the control group students receive less quiz feedback than experimental students.

Four variables relevant to the present study described by Kulik et al. (1990) as procedures used in Mastery Testing seem appropriate to consider:

1. *Pacing*. Students in the Mastery Learning programs proceeded through a course at their own pace or progressed through material as a group.
2. *Mastery level on unit Tests*. Programs varied in the percentage correct needed to establish mastery on a unit test.
3. *Demonstration of mastery*. Some programs required a formal demonstration of mastery on each unit test (i.e., students had to take alternative forms for unit tests until they reached a prespecified mastery level of performance), whereas in other programs mastery could be demonstrated less formally by completion of prescribed remedial activities.
4. *Duration of treatment*. Programs varied in the number of weeks of duration.

In addition, the following seven variables used to describe the experimental designs of the studies are applicable to the present study:

1. *Subject assignment.* Students were assigned to experimental and control groups either randomly or by nonrandom procedures.
2. *Teacher effects.* In some studies the same instructor or instructors taught both experimental and control groups, whereas in other studies different instructors taught experimental and control groups.
3. *Historical effects.* In some studies experimental and control groups were taught concurrently (e.g., in the same semester), whereas in other studies experimental and control groups were taught consecutively (e.g., in two different semesters).
4. *Frequency of testing.* In some studies experimental and control groups took the same number of unit tests. In other studies students in the control group were tested less frequently than students in the experimental group.
5. *Amount of quiz-feedback.* In some studies experimental and control group received the same amount of feedback on unit quizzes. In other studies, however, amount of feedback for experimental and control students differed for one of two reasons: a) control students took fewer quizzes than did experimental-group students and thus necessarily received less feedback, or b) experimental and control students took the same number of unit quizzes but experimental-group students received only information on total quiz scores.
6. *Locally developed versus standardized criterion tests.* Studies used either local tests, nationally standardized tests, or a combination of the two.
7. *Objectively versus subjectively scored criterion tests.* Some studies used objective, machine-scoreable criterion examinations, whereas others used essay tests or other nonobjective tests to measure final performance.

Table 9.0

Summary of Kulik and Kulik, Banger-Downs (1990)
Analysis of Characteristics Related to the Present Study:

<u>Procedure variables:</u>		
	Phase I	Phase II
Pacing	group	grp./indiv.
Mastery level on unit tests	85%	85%
Demonstration of mastery	formal	formal
Duration of treatment	4 weeks	4 weeks
<u>Experimental variables:</u>		
	Phase I	Phase II
Subject assignment	nonrandom	nonrandom
Teacher effects	different	n/a
Historical effects	concurrent	n/a
Frequency of testing	equal no. unit test	n/a
Amount of quiz feedback	treat. feedback none for control	treat. feedback
Locally developed vs standardized criterion tests	locally developed	locally developed
Objectively vs subjectively scored criterion tests	both	both

As can be seen by Table 9.0, the present study's Phase I contains more features that are applicable to the meta-analysis studies since the design (Design 10, Stanley and Campbell, 1963) included a control group. Although Phase II of the study is more difficult to include, it does follow the procedure variables identified.

Kulik and Kulik et al., have concluded that the data in their analysis suggest that effects of Mastery programs are not uniform on all students in class, that perhaps low aptitude students gain more. Variation on final exam scores is smaller in Mastery classes than in other classes. In addition, the correlation between aptitude and achievement is reduced in the Mastery classes. Finally there appears to be relatively enduring not just short term effects and also influences exam performance which has positive effect on student

attitudes. Because Phase I (the only design that fits inclusion in this analysis) was poorly implemented, not one of these findings could even be claimed for the Mastery Learning treatment in Phase I.

Another recent scrutiny of Mastery Learning research has been conducted by Slavin (1987). Slavin's "best evidence synthesis" provides two areas or dimensions of challenge for the present study: unequal time and unequal objectives. This analysis also includes some of the features scrutinized by Kulik et al. although Kulik et al. included in their analysis more features and a wider spectrum of studies.

Slavin claims unequal time occurs when total instructional time allocated to a particular subject is fixed, then a common level of learning for all students is likely to require taking time away from high achievers to increase it for low achievers. Thus, he claims that many Mastery Learning programs provide corrective instruction during times other than regular class time. According to Slavin, in many of Bloom's studies corrective instruction is given outside of regular class time, thus increasing total instructional time beyond that allocated to the control groups. Studies that fail to hold time constant across treatments essentially confound treatment effects of additional time. On the other hand, when time for corrective instruction is provided during regular class time (rather than after class or after school), Mastery Learning trades coverage for time.

Unequal objectives Slavin states, correspond more closely to the curriculum taught in Mastery Learning experiments. Mastery classes spend time on mastering a limited set of objectives while the control group may have learned a larger set of objectives (though perhaps at lower level of mastery). There is a danger in using Formative and Summative Tests. Many studies administer the Formative and Summative Tests used in Mastery Learning classes as quizzes in the control; theoretically it should help focus the control classes on the same objectives as the mastery classes especially, if the

same texts and materials are used. When careful control of instruction methods, materials, and tests is not exercised there is always a possibility that the control group is learning valuable information or skills not learned in the Mastery Learning group but not assessed on experimenter made measures. Essentially, Slavin concludes that the possibility that even though all teachers used the same materials, the Mastery Learning teachers focused on the specific objectives to be tested more than the control classes did.

Again, as with the Kulik et al. analysis, the issues raised by Slavin only apply to Phase I and not Phase II of this study. Phase I included a control group for comparison. The issue of unequal time is not applicable since no effort was made to schedule any extra time outside of class time for correctives and/or feedback for practice or remediality. The issue of unequal objectives is an open question. While the treatment group regressed on most dependent measures there was a concerted effort to adhere to the same objectives in both the control and treatment classrooms. There was also an effort made in using the same text and same instructional material even though the treatment group was exposed to supplemental material that the control group did not get. The treatment implementation did not include all aspects of the supplemental material that was originally planned, however.

Conclusions and Interpretations

Without a thorough analysis of the strengths and weaknesses of the study's methodology it was very difficult to draw any conclusions. This study had an 'extra burden of proof, in that instead of using only one design, it utilized two distinct designs. The implementation of the treatment in the first phase of the study raised particular issues. These circumstances made it necessary to seek an

alternative site for a more favorable condition in order to test whether, in fact, a Mastery Learning approach was viable for teaching basic writing skills to bilingual fourth graders. Hence, Phase II of the study was implemented.

According to an analysis of factors in the research methodology literature, the first phase of the study has a stronger research design since, if implemented properly, this design controls for various extraneous threat to the internal validity (see section on internal validity Chapter VI). While Phase I of the study has a stronger design, the sloppy implementation of the treatment actually weakened the internal validity of the findings. The second phase of the study used a design that is characterized as a pre-experimental design thus it lacks internal validity. Design 2 Phase II does not have a control group for comparison.

It seems that the designs for both phases are relatively strong in terms of external validity according to Snow's Model of quasi-Representativeness. Generally, it can be surmised that the relationships identified in the study represent "real" phenomena in the sample of students that participated in the study. Therefore, based on quasi-representativeness, it seems reasonable to state with some level of confidence that the findings are generalizable beyond the sample used for the study.

The pattern of the findings yielded by the analysis allow three basic conclusions:

1. The Mastery Learning approach is a viable means for teaching basic writing skills in a classroom composed of diverse language proficiency skills.
2. Achievement can and does increase when program implementation conditions are favorable. Learning basic writing skills at the fourth-grade level is crucial for further learning, and a

Mastery Learning approach makes instruction manageable in a classroom where it is of utmost importance to meet diverse needs of linguistically and culturally diverse pupils. There was considerable progress in achievement in the well implemented Mastery Learning classroom. On the other hand, the classrooms used for the study in Phase I show regressive results for the treatment classroom and inadequate acquisition of main idea paragraph writing skills in the control classroom.

3. The key variables in determining proper implementation of equity education programs are teacher beliefs and judgment. In Phase I of the study, the practice of tracking into ability sections clearly diminished academic achievement and promoted negative interactions in the classroom environment which in turn can be seen as contributing to emotional consequences for students' academic self-perceptions. In Phase II there was no tracking into ability sections, students were placed within the same classroom in instructional groups not according to ability, but rather based on linguistic proficiency. Bilingual students could choose either language (Spanish or English). The interactions in the classroom were generally positive.

Each of these conclusions are interrelated phenomenon and will be expanded in the following section.

Conclusion One

This study has tested achievement through an approach for teaching basic writing skills regardless of the language of instruction. In addition, the writing samples gathered as the result of an instructional measure were assessed by various dimensions of writing. It has been shown in previous studies (i.e., Duran, 1983)

that the acquisition of writing skills of bilinguals is not an easily understood phenomenon. It is clear that Hispanic bilingual students do not develop the level of academic writing skills that are necessary for academic success (in either Spanish or English). The question of assessment and appropriate instruction of the writing process rather than the final written product oriented assessment and instruction is of major concern to educators.

The results of this study clearly show that if instruction is fragmented, subjectively evaluated, and not closely monitored, it can have a detrimental effect on the achievement and progress of the students. The Mastery Learning treatment in Phase II of this study was closely monitored and such detrimental effects seemed less evident. Individual students were given additional practice and encouraged on an individual basis to write more and feedback was consistently provided. The close monitoring provided information about the skill acquisition of monolingual (English or Spanish) or bilingual students and thus instructional decisions were made on an ongoing basis. This continual activity provided an equitable learning environment for the students, and, in addition, allowed the teacher to understand how much additional time was needed for learning for the various levels of basic linguistic ability that existed in one classroom. The results of Phase II of the study indicate that, indeed, a specific set of tasks were sequentially followed and monitored for learning to take place.

It seems logical to propose that a well implemented Mastery Learning approach can be used to provide a sound basic skills writing program. Such a program can allow writing to be evaluated and monitored as a process rather than only as products that students produce on a fragmented basis.

Conclusion Two

Because Mastery Learning uses a series of criterion-reference measures to monitor student progress, it is often wrongly concluded that it is not a usable approach for teaching writing skills. The premise taken by this study is that writing tasks are not only grammar oriented or mechanics but more importantly and ideally they are "idea formation exercises" which enhance cognition. Additionally, abstract thought, taught through learning to write, is an essential skill that all students at the fourth-grade level must learn in order to have academic success. It is clear, that in Phase I of the study, instruction in writing for both the control and the treatment classroom was fragmented and that there was little growth or no growth or progress in main idea acquisition. Phase II, on the other hand, reveals a very positive growth pattern on all dimensions of writing.

It is very clear then, that the organizing, monitoring, and sequencing of skills allows the teacher not only to gauge the elements of writing that are typically taught and presumably can be gauged periodically by criterion-reference measures, but also this "harnessing" of the basic levels may 'free up' time, energy, and resources to teach. This will also provide learning opportunities for language elaboration and creative language use ,thus promoting and stimulating abstract thought through language. This approach was successful in the Phase II treatment

Conclusion Three

The literature review also discussed the negative effects of competitive environment and ability tracking based on subjective evaluations of teachers. The key ingredient in any equity learning

program is teachers' beliefs and subjective judgments (Block et al., 1989). It has been hypothesized that if students are not perceived by their teachers as competent, they begin to develop mechanisms to save and protect their ego and their performance and motivation are hindered.

This study attempted to gauge affect in various ways, but the achievement results clearly revealed that tracking was detrimental for the low ability classroom (which was also the treatment classroom) in Phase I. The students in this low ability classroom regressed on virtually all measures despite being approximately equal in skill at Pretest level to the higher ability class.

What are the reasons for these results? It is certain that the Mastery treatment was improperly implemented. We suspect that the sort of implementation problems stemmed from the teachers subjective evaluation of writing ability. Limited English proficient students were seen as 'language handicapped' and this led the low ability, Mastery classroom to be poorly managed. The evidence presented in this study shows severe regressive results on all skills dealt with during a four week instructional sequence. The teacher continually compared her students to a norm and not to what they were learning on a daily basis. The results of their daily performance were not used to modify nor create additional materials, lesson presentations, or monitoring systems so that they therefore, could not succeed.

Implications for Future Research

Despite the fact that bilingual education programs have been implemented in the public schools for approximately twenty years there is a paucity of research and theory on effective instructional programs for teaching bilingual or monolingual Spanish language background students. Research related to meeting the educational needs of Mexican American students on methods and curriculum that utilized experimental methodology is virtually non-existent.

Critical issues facing U. S. education have been identified by Chicano education experts as changing demographics, changing world economy, and the creation of an educational system that is school-work relevant. In addition to other elements outlined for an effective school model, Leticia Quezada (the only Hispanic on the seven member Los Angeles School Board) identified as an educational goal, the notion that all teachers in the same school have the awareness of and work toward the same educational goals. Also, these goals include constant evaluation of objectives, daily, weekly, monthly and yearly, are seen as priorities for effective public education (Hispanic Weekly, 9/17/90). In order to work towards these concerns we need to focus research experiments to test out equity educational models with Hispanic students as subjects. The replication of studies such as the present study with different groups of Chicano students from different communities, SES backgrounds, and levels of bilingual proficiency linguistic skills, will provide knowledge concerning within group differences which in many research projects in the past have been glossed over.

Though more than two decades of research has specified reasons for the nations high dropout rate and leaving school early, especially for minority students, not enough research has been conducted that has strong bearing on sound educational policy.

According to Block et al. (1989), public policymakers are taking a hard look at issues of equity in student learning. These reformers it seems advocate equality of educational outcome as well. What is sorely needed are experiments that take into consideration students' attitudes towards learning especially minority students' attitudes. In addition, more studies need to be conducted that take into account not only within group factors (diversity of Mexican American community), but how these factors are manifested in the classroom in the form of teacher attitudes, curriculum, modes of instruction, and motivational structures for learning. In sum, the existent research is not sufficient to get a clear picture of institutional factors that promote pushing minority students out of school early. The focus of much of the research that has been done in the past on Mexican American students is on linguistic issues at the expense of other factors that tear down self-esteem and a strong positive perception of achievement.

Language development is a 'live' ongoing process that is influenced by many factors in the student's environment; it is dynamic. Therefore, it is crucial that language programs designed to meet the needs of the diverse linguistic proficiencies of Spanish language background students are relevant to the daily lives of students. Furthermore, these language processes have inherent observable phenomenon as well as non-observable phenomenon that cannot be always be determined by linguistic output. Thus it is important to evaluate, gauge, modify, and plan relevant activities on a continual basis to insure learning and promote a good perception of achievement linkage.

The content of this Mastery Learning study was the teaching of basic writing skills. If fourth-grade bilingual children do not acquire the necessary basic writing skills by the end of the fourth-grade it is very unlikely that they will be academically successful. There are many aspects of linguistic development that seem to be interacting at

the same time in a classroom setting. The theory of writing proposed by the literature in this study is that writing is a form of communication and that it is manifested in the classroom as a social activity. This learning activity is crucial for the development of cognitive language skills through writing. When we consider the complexity of abstract thought and how it contributes to language output, it is important to design instructional programs that are viable in meeting the needs of different levels of linguistic proficiency in all students, especially in the students that are developing bilingual language skills (the use of two languages simultaneously, in parallel, or sequential development).

According to Freeman and Pringle (1979), for children learning to write, the transfer from oral skills to the written medium occurs early in life when competency in the oral medium of the native language is still far from fully established, when many years of physical, intellectual, cognitive, emotional and moral maturation still lie ahead. These various aspects of development, these researchers contend, obviously interact in complex ways with the acquisition and development of writing abilities.

Bloom (1964) has extensively researched how intelligence is viewed in educational settings and concludes that verbal ability represents a very important part of most general intelligence. The notion of intelligence as being a construct that is indeed a developing phenomenon, is extremely important to consider when developing educational programs for bilingual and bicultural students. In most educational programs achievement and hence intelligence is measured through verbal tasks. In the past, bilingual Mexican American/Chicano children have been continually diagnosed as learning disabled based on instruments used to test intelligence in the public schools. One can venture to say that presently this practice still exists, only it takes a different form in that many children are

simply tracked into ability groups and remain in these groups throughout their schooling.

Generally, the school's personnel feel that children that speak another language should start school with a level of formal language development necessary to succeed in academic settings, but this is simply not the reality. School programs that are designed to transition these students into the second language are doing it at the expense of their cognitive development. What seems to be occurring to these bilingual, bicultural students is confusion. Programs such as those designed, like the ones used in Phase I of this study, rather than promote learning and achievement may be more detrimental to students' academic development and hence their self esteem. While one can observe that these students make a tremendous effort at assimilating into the mainstream (in terms of behavior, language preference, and language use) basic language skills necessary for academic success are not developed in either language (Spanish or English) by a great number of Spanish language background students. Instead, second language structures are developed at a very superficial level of ability. Second language programs in many schools are designed for fast, functional acquisition of language for survival purposes. It has been observed that if students seem to use the second language orally then it is assumed that he or she is capable of continuing in an English-only curriculum. This evaluation is done without consideration for deeper level cognitive development, abstract thought process, and stronger comprehension of the second language. What appears to happen is that it is assumed that these students are not capable of higher level learning and very early on in their academic careers they are channelled and tracked into low-ability classrooms.

Psychological factors discussed previously begin to manifest in very complex ways in the context of the classroom. As substantiated by the literature, the ways in which these constructs are manifested

are not yet clearly understood by educators or researchers. Within the Spanish language background group of students there are different cultural backgrounds, economic classes, assimilation patterns, and learning styles. All of these factors taken together promote different types of attributions for success or failure (see Weiner, Chapter II).

Bloom (1964) supports the view that environments which include good models of language will encourage development of general intelligence, whereas environments in which models of language are poor will discourage language development, and furthermore will retard or block the development of general intelligence. The language environment of the classroom is perhaps the most important environment for the development of the formal language of schooling and, hence, general intelligence. In accordance with Block et al., (1989) the important thing is not just particular learning styles, even though teachers should understand the diversity that exists in the Spanish language background group population, but the learning outcomes and the means for insuring that the outcomes are positive for all and not just a few.

The classroom environment of the conventional classroom has failed to meet the needs of students of Spanish language background. The high dropout rate of these students continues to increase. The effects of negative school experiences becomes an accepted struggle that does and will continue to affect students throughout their school experience. As can be seen from the results of this study, these negative school practices start early. These school programs are not meeting the needs of these students because they are not designed as equity structures.

REFERENCES

- Alexander, K. L., Fennessey, J., McDill, E. L., and D'Amico, R. J. (1979). School SES influences: Composition or context? *Sociology of Education*, 52, 222-237.
- Ames, C. & Ames R. (1981). Competitive versus individualistic goal structures: The salience of past performance information for causal attributions and affect. *Journal of Educational Psychology* Vol.73, No. 3 411-418.
- Ames, C. (1984). Competitive, cooperative, and individualistic goal structures: A cognitive-motivational analysis. In Russell Ames and Carole Ames (Eds.) *Research on Motivation in Education Student Motivation* Volume 1 Academic Press.
- Ames, C., and Felker, D. W. (1979). Effects of self-concept on children's causal attributions and self reinforcement *Journal of Educational Psychology*, 71, 613-619.
- Anania, J. (1983). The influence of instructional conditions on student learning and achievement. *Evaluation in Education: An International Review Series*, 7, 3-76.
- Anastasi, A. (1964). Some current developments in the measurement and interpretation on test validity. In Lewis R. Aiken Jr. (Ed.) *Readings in Psychological and Educational Testing*. Allyn and Bacon, Inc. Boston.
- Baker, Eva (1982) The specification of writing tasks. *Evaluation in Education* Vol. 5 pp. 291-297.
- Beck, L., Muia, J. (1980) A portrait of a tragedy: Research findings on the dropout *The High School Journal* The University of North Carolina Press.
- Block, J.H. (1970). *The effects of various levels of performance on selected, cognitive, affective, and time variables*. Unpublished doctoral dissertation, University of Chicago.

- Block, J. H. (1974). Mastery learning in the classroom: An overview of recent research. In J. Block (Ed.), *School, society, and mastery learning*. New York: Holt, Rinehart, and Winston.
- Block, J. H. (1977). Motivation, evaluation, and mastery learning. *UCLA Educator*, 19, 31-36
- Block, J. H., & Burns, R. B. (1976). Mastery learning. In L. Shulman (Ed.), *Review of research in education* (Vol. 4). Itasca, IL: Peacock.
- Block, J.H., Efthim, H.E., & Burns, R.B. (1989) *Building Effective Mastery Learning Schools*. Longman, New York, London.
- Bloom, B.S. (1970). Towards a theory of testing which includes measurement-evaluation-assessment. In M.C. Wittrock and D.E. Wiley (Eds.) *The evaluation of instruction*. New York: Holt, Rinehart and Winston.
- Bloom, B.S. (1971). Affective consequences of school achievement. In J.H. Block (Ed.), *Mastery learning: Theory and practice*. New York: Holt, Rinehart, and Winston.
- Bloom, B.S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill
- Bracht, G.H. & Glass, G.V. (1968) The external validity of experiments. *American Educational Research Journal*. Vol. 5, No.4.
- Bricklen, B., Bricklen, P.M. (1967). *Bright Child-Poor Grades..* New York: Dell.
- Brookover, W. B. (1965). *Self Concept of Ability and School Achievement*. Bureau of Educational Research Services College of Education Michigan State University, East Lansing Michigan.
- Brophy, J.E., (1976) Reflections on research in elementary schools. *Journal of Teacher Education* Vol. 27, No. 1.

- Brophy, J.E. & Good, T.L. (1970) Teachers communication of differential expectations for children's classroom performance: Some behavioral data. *Journal of Educational Psychology*, Vol.61, No. 5, 365-374.
- California Department of Education (1985). *Dropping Out, Losing Out: The High Cost for California* Assemble Office of Research
- California School Boards Association (1985). *Make Something Happen. Hispanics and Urban High School Reform* National Commission on Secondary Schooling for Hispanics Volume I.
- Camp, C. (1980). *School Dropouts A Discussion Paper*. California Legislature Assembly Office of Research.
- Campbell, D. T., Stanley J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.) *Handbook of Research on Teaching* . Rand McNally and Company Chicago.
- Campbell, D.T.,and Fiske,D.W. (1956). Convergent and discriminant validation by the multitrait-multimethod matrix *Psychological Bulletin* 56, 81-105.
- Carter, T. & Segura R. (1979). *Mexican Americans in School A Decade of Change* . New York: College Entrance Examination Board.
- Casas, J., and Ponterotto, J. (1984). Profiling an invisible minority in higher education: The chicana. *Personnel and Guidance Journal*, 62, 349-353.
- Casas, M, J., Furlong M. J., Carranza, O. & Solberg, S. (1986). *Santa Barbara Student Success Study: Profiling Successful and At-Risk Junior High School Students*. Final Report Submitted to the Santa Barbara School Districts Board of Education.
- Casas, M. J., Furlong M. (1986). In search of an understanding and a responsible resolution to the Mexican-American educational dropout problem. *California Schools Forum* Vol. I Graduate School of Education University of California Santa Barbara, California.

- Covington, M. V. (1984) The motive for self-worth. In Russell Ames and Carole Ames (Eds.) *Research on Motivation in Education Student Motivation Volume 1* Academic Press, Inc.
- Covington, M.V. and Omelich, C.L. (1979). Effort: The double-edge sword in school achievement. *Journal of Educational Psychology*, 71, 169-182.
- Covington, M. V., Omelich C. L. (1985). Ability and effort valuation among failure-avoiding and failure-accepting students. *Journal of Educational Psychology*, Vol. 77, No. 4, 446-459.
- Covington, M. V., Spratt, M. F., Omelich, C. L. (1986). Is effort enough, or does diligence count too? Student and teacher reactions to effort stability in failure. *Journal of Educational Psychology*, Vol. 72, No. 6 717-729.
- Darehshori, C. (1977) The way it is. *Today's Education* March-April.
- DeAvila, E. A., B. Havassy, J. Pascual-Leone (1976). *Mexican American School Children: A Neopagetian Analysis*. Washington, DC: Georgetown University Press.
- DeAvila, E.A. and Duncan S. E. (1981). Bilingualism and the metaset. In Richard Duran (Ed.) *Latino Language and Communicative Behavior*. Norwood New Jersey.
- deCharms, R. (1968). *Personal causation*. New York: Academic Press.
- deCharms, R. (1984). Motivation enhancement in educational settings. In Russell Ames and Carole Ames (Eds.) *Research on Motivation in Education Student Motivation Volume 1* Academic Press, Inc.
- Deci, E. L. (1975). *Intrinsic motivation*. New York: Plenum.
- Deci, E. L. (1980). *The Psychology of Self-Determination*. Lexington Books D.C. Heath and Co.
- Dentler, R. A. & Warshauer, M.E. (1968). *Big city dropouts and illiterates*. New York: F.A. Praeger Publishers.

- Duran, R. P. (1983). *Hispanics' Education and Background Predictors of College Achievement* College Entrance Examination Board, New York.
- Duran, R. P. (1986). Testing of linguistic minorities.
In R. Linn (Ed.) *Educational Measurement* New York: McMillan.
- Engle B.S. (1977) One way it can be. *Today's Education* March-April.
- Erickson, F. (1987). Transformation and school success: The politics and culture of educational achievement.
Anthropology and Education Quarterly Vol. 18 pp. 335-355.
- Ford Foundation (1984). *Hispanics: Challenges and Opportunities*. New York: Ford Foundation.
- Freeman, A & Pringle I (1979) *Learning to write: first language/second language*. New York: Longman.
- Galloway, D. (1985). *Schools and persistent absentees*. Oxford: Pergamon Press.
- Guskey, T. R., & Gates, S. L. (1986). Synthesis of research on the effects of mastery learning in elementary and secondary classrooms. *Educational Leadership*, 43(8), 73-80.
- Harrari, O., and Covington M.V., (1981). Reactions to achievement behavior from a teacher and student perspective: A developmental analysis. *American Educational Research Journal*, 18, 15-28.
- Hernandez, N. G. (1973). Variables affecting achievement of middle school Mexican-American students. *Review of Educational Research* Vol. 43, No. 1, pp. 1-39.
- Johnson, J. N. (1975). *Business Council for Effective Literacy, Adults in Crisis, an Adult Performance Level Study*, University of Texas, Arlington Texas.
- Kash, M. M., Broich, G. D. (1982). Teachers. In Herbert J Walberg (Ed.) *Improving educational standards and productivity The research basis for policy*. McCutchen Publishing Corp.

- Kozol, J. (1985). Why 60 million U.S. adults wouldn't be able to read this," *Sacramento Bee* Sacramento California.
- Kulik, C.L.C., Kulik, J.A. & Bangert-Downs, R.L. (1990) Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research*. Vol. 60, No. 2.
- Kulik, C.L.C., Kulik, J.A. & Bangert-Downs, R.L. (1990) Is there better evidence on mastery learning? A response to Slavin. *Review of Educational Research*. Vol. 60, No. 2.
- Lane, M. A. (1984). *State Library Literacy Campaign Program Effectiveness Review*. Submitted to the California State Library, Sacramento, California Department of Education.
- ✓ Laosa, L. M. (1977). Cognitive styles and learning strategies research: Some of the areas in which psychology can contribute to personalized instruction in multicultural education. *Journal of Teacher Education*, 28, 26-30.
- ✓ Laosa, L. M. (1977). Inequality in the classroom: Observational research on teacher-student interactions. *Aztlan: International Journal of Chicano Studies Research*, 8 51-67.
- ✓ Laosa, L. M. (1984). School, occupation, culture and family: The impact of schooling on the parent-child relationship. *Journal of Educational Psychology*, 74, 791-827.
- Lloyd-Jones, R. (1977). Primary trait scoring. In Charles R. Cooper and Lee Odell (Eds.) *Evaluating Writing: Describing, Measuring, Judging* State University of New York at Buffalo.
- Los Angeles Unified School District. (1985). *A Study of student dropout in the Los Angeles unified school district*. Los Angeles.
- McDermott, R.P. (1987) The explanation of minority school failure, again. *Anthropology and Education Quarterly* Vol. 18 361-364.
- McKenna, B. (1977) What's wrong with standardized testing? *Today's Education* March-April
- Moll, L. C., Diaz S. (1987). Change as the goal of educational research. *Anthropology and Education Quarterly* Vol. 18.

- Nicholls, J. (1984). Conceptions of ability and achievement motivation. In Russell Ames and Carole Ames (Eds.) *Research on Motivation in Education Student Motivation Volume 1* Academic Press, Inc.
- Nicholls, J. G. (1975). Causal attributions and other achievement-related cognitions; effects of task outcome, attainment value and sex. *Journal of Personality and Social Psychology* 31, 379-389.
- Nicholls, J.G. (1976). Effort is virtuous, but it's better to have ability: Evaluative responses to perceptions of effort and ability. *Journal of Research in Personality*, 10, 306-315.
- Nicholls, J.G. (1978). The development of the concepts of effort understanding that difficult tasks require more ability. *Child Development*, 49, 800-814.
- Nicholson, S. J. (1988) *Motivational Aspects of Using Computers to Teach Writing*. Unpublished Doctoral Dissertation University of California at Santa Barbara.
- Nordin, A.B. (1979). *The effects of different qualities of instruction on selected cognitive, affective, and time variables*. Unpublished doctoral dissertation, University of Chicago.
- Oakes J. (1984). Tracking and ability grouping in American schools: Some constitutional questions. *Teachers College Record* Vol. 84, No. 4.
- Ogbu, J. (1978). *Minority Education and Caste: The American system in cross cultural perspective* New York: Academic Press.
- Ogbu, J. (1981). Origins of human competence: A cultural ecological perspective. *Child Development*, 52, 413-429.
- Ogbu, J. (1982). Cultural discontinuities and schooling. *Anthropology and Education Quarterly* Vol. XIII, No. 4
- Orozco-Suarez, M. (1987). Towards a psychosocial understanding of Hispanic adaptation to American schooling. In H. Trueba (Ed.) *Success or Failure the Learning and Language of Minority Students* Newbury House.

- Polin, L. (1985). *The Effect of Strategy Assistance on the Writing Performance of Tenth Grade Students: A Test of the "Writer Overload" Hypothesis*. Unpublished Doctoral Dissertation University of California at Los Angeles.
- Quinto, F. (1977) Teacher-made tests-an alternative to standarized tests. *Today's Education* March-April.
- Rist, R. C. (1970). Student social class and teacher expectations: The self fulfilling prophecy in ghetto education. *Harvard Educational Review*, 40.
- Rodriguez, J. H. (1985). *Toward Effective Education for Hispanic Youth A Leadership Challenge* San Diego County Office of Education.
- Rosenbaum, J. E. (1976). *Making inequality: The hidden curriculum of high school tracking*. New York, NY. John Wiley and Sons.
- Rosenbaum, J. E. (1978). Structure of opportunity in schools *Social Forces*, 57 (1) , 236-256.
- Rumberger, R. W. (1982). Recent high school and college experiences of youth: Variations by race, sex, and social class. *Youth and Society* Vol.13, No. 4. Sage Publications Inc.
- Rumberger, R. W. (1983). Dropping out of high school: The influence of race, sex, and family background. *American Educational Research Journal*, Vol. 20, No.2 pp. 199-220.
- Rumberger, R.W. (1987) High school dropouts: A review of issues and evidence *Review of Educational Research* Vol. 57, No. 2, pp.101-121
- San Diego City Schools. (1985). *The 1982-83 leaver study of the San Diego unified school district..* Planning Research and Evaluation Division. San Diego California.
- San Miguel, G. (1986). Cycles of concern: A historical perspective on the dropout problem. *California Schools Forum*, Vol. I. Graduate School of Education University of California, Santa Barbara California.

- Sanchez, G.I. (1932). Group differences and Spanish-speaking children: A critical review *Journal of Applied Psychology*, 16, 549-558.
- Sanchez, G.I. (1934). Bilingualism and mental measures: A word of caution. *Journal of Applied Psychology*, 8, 765-772.
- Schrieber, D. (1964). The school dropout-A profile. *Educational Digest*, 62, 229-243.
- Seligman, M.E.P. (1975). *Helplessness*. San Francisco: Freeman.
- Slavin, R. E. (1987). Mastery learning reconsidered. *Review of Educational Research*, Vol. 57, No. 2, pp. 175-213.
- Slavin, R. E. (1990). Mastery learning re-reconsidered. *Review of Educational Research*, Vol. 60, No. 2, pp. 300-302.
- Snow, R.E. (1974) Representative and quasi-representative designs for research on teaching. *Review of Educational Research*. Vol. 44, No. 3.
- Sohn, D. (1977). Affect-generating powers of effort and ability self attributions of academic success and failure. *Journal of Educational Psychology*, 69, 500-505.
- Steingberg, L. Blinde P., Chan K. (1984). Dropping out among language minority youth. *Review of Educational Research* Vol. 54, No. 1 pp. 113-132.
- Taylor, E.F. (1977) The looking-glass world of testing. *Today's Education* March-April.
- Tinsey, H.E.A. & D.J. Tinsey, (1987). Uses of factor analysis in counseling psychology research. *Journal of Counseling Psychology*, Vol. 34, No. 4, 414-424.
- Treadway, P. G. (1985). *Beyond Statistics: Doing something about dropping out of school A synthesis of relevant research. Whatever will happen to the class of 1992?* Whitney Education Foundation School Drop-Out Prevention Conference Cupertino California.

Walker, C. (1987). Hispanic achievement: Old views and new perspectives. In Henry Trueba (Ed.) *Success or Failure the Learning and Language of Minority Students* Newbury House.

✓ Wehlage, G. G., Rutter, R. A. (1986). Dropping out: How much do schools contribute to the problem? *Teachers College Record* Vol. 87 No. 3.

Weiner, B. (1979). A theory of motivation for some classroom experiences. *Journal of Educational Psychology*, 71,3-25.

Weiner, B. (1984). Principles for a theory of student motivation and their application within an attributional framework. In Russell Ames and Carole Ames (Eds.) *Research on Motivation in Education Student Motivation Volume 1* Academic Press, Inc.

White, R. (1959). Motivation reconsidered: The concept of competence. *Psychological Review* 66, 297-333.

Appendices:

Appendix A

1. Table of Specifications, testing/
instructional materials Phase I
2. Table of Specifications, testing/
instructional materials Phase II

Appendix B

1. Sample Affect Questionnaire

Appendix C

1. CSE Subscales and Trait Scales Scoring Rubrics

Appendix A

Phase I

1. Table of Specifications
2. Instructional Materials: Formative and Summative Tests

Phase II

1. Table of Specifications
2. Instructional Materials: Formative and Summative Tests

TABLE OF SPECIFICATIONS : Phase I

Basic Textbook: Language Skills and Use by Scott, Foresman and Co.

ObjectivesStudent Behavior

	know	comp	apply	anal	synth	eval
Part 1.1						
What is a Sentence?						
-recognizes what a sentence is	*	*	*			
-identifies abstract words tused to describe an idea	*	*	*			
-uses basic senses/emotions to make up words for sentences	*	*	*			
Part 1.2						
Grouping Sentences/Punctuation						
Capital letters						
-recognizes / uses capital letter lbeginning a sentence	*	*	*	*		
-identifies period as punct. mark	*	*	*	*		
Part 1.3						
Kinds of						
Sentences/Punctuation Marks						
-recognizes sentences with three different punctuation marks	*	*				
-identifies use of period in sentence	*	*	*			
-identifies question mark in sentence	*	*	*			
-identifies exclamation marks in a sentence	*	*	*			
-recognizes t differences between three punctuation marks	*	*	*	*		
-recognizes different kinds of sentences depending ending mark	*	*	*	*	*	
Part 1.4						
Writing Paragraphs						
-groups together sentences with same idea	*	*	*	*		
-recognizes a paragraph	*	*				
-identifies use of indentation	*	*	*			
-identifies basic characteristics of a paragraph;ie.one main idea unified and non-fragmented coherent , continuous organization	*	*	*	*		
-uses details for descriptive para graph	*	*	*			
-recognizes sentence with single idea	*	*	*			
-differentiates between a complete and fragmented sentence	*	*	*			

Unit Lesson Plans		Sentence and Paragraph Writing	
Week	Aspect of Teaching	Activities	Grp/Mode
1.	A. Orientation B. Presentation	1. Class discussion 1. students aware of need for skill 2. describe an action or idea that is abstract 3. write words relating to senses or emotions 4. write sentences using words	whole grp whole grp. whole grp.
2.	A. Orientation B. Presentation Text Part 1.1 p. 30-31 grp. C. Involvement	1. List reasons for need of skill 1. Recognize sentences, words together to clarify idea 1. make up own sentences	whole grp. whole grp. seat work indiv. individual/pair whole grp. whole grp.
2.	A. Orientation What is sentence? B. Presentation C. Involvement Part. 1.2 Text p. 32,33,36,37	1. Discuss topic 1. grouping sentences 1. The use of capital letters 2. the use of a period 3. the use of punctuation	seat whole grp.
3.	A. Orientation B. Presentation part 1.3 pg. 40-41 C. Involvement	1. Discuss need for skill 1. Text content kinds of sentences 2. generate own sentences 1. recognizes diff. types of punctuation marks and sentences 2. writes sentences using marks	whole grp. whole grp. indiv/ whole grp. indiv./pair
4.	A. Orientation B. Presentation whole grp. Part 1.5 pg.90-91 C. Involvement	1. List reasons for need of skill discuss 1. recognize shape, length, function of paragraph 2. introduction to indentation 3. state essential characteristics of paragraph 4. select topic sentence 5. summarize rules 1. recognize orderly and disorderly paragraphs 2. group sentences together to form paragraphs 3. indent and use capital letters appropriately 4. use punctuation correctly	whole grp. indiv. indiv./pairs

Formative Test I

A. Circle the words below that are words which can be used to describe how you feel or how things look, feel, smell, taste or sound.

desk	house	pencil	car
soft	sour	happy	hard
tree	rain	building	sad
sloppy	delicious	pizza	neat
loud	hot	clear	dark

B. Circle each group of words below which are the ones that tell you something clearly.

1. I think does nice.
2. It broken came loud.
3. Joe is from another neighborhood.
4. Is she a pretty girl?
5. The banana smells delicious!
6. There are three school buses.
7. Hot dog people green.
8. Skipper is Bob's dog.
9. The leaves on the trees are dry
10. Ice-Cream will taste cold and smooth.

Formative Test 2 (week one)

A. Circle the words below that are words that can be used to describe how you feel or how things look, feel, smell, taste or sound.

hat	nice	dark	cup
card	hammer	table	book
pen	honest	hard	round
messy	chair	paper	colorful
plain	plate	fork	pin

B. Circle each group of words below that tell you something clearly.

1. John table fell.
2. Luz jumped on one foot.
3. Doug fell off his bike.
4. Father book chair in room is.
5. She has a big box of books.
6. Halloween good costumes for.
7. The teacher is going trick or treat for Halloween.
8. She does put foot.
9. He asked his mother for a nice Halloween costume.
10. Today was a sunny and beautiful day.

Correctives week 1- Moments book

B. Animals

Names of things in the *animal* group

Words that tell what all animals can do

D. Living things

Write the names of five living things you can see on pages iv and 1 of the Moments book. Then list some words that tell what each living thing can do.

Living things What they can do

Corrective- week 1

Some things I can taste:

[illegible]

Some things I can smell:

Some things I know by using more than one sense:

Thing

Sense

[illegible]

Correctives week 1

A. Dogs

A picture of the thing we all agree to call dog:

What all (or most) dogs can do:

Formative Test 1 Week 2 (Capt. and Punct.)

A. Circle the letter of the ending mark that should come at the end of each sentence.

1. That lamp is falling

a. . b. ? c. !

2. It looks cloudy

a. . b. ? c. !

3. When is the party

a. . b. ? c. !

4. The stamps are green

a. . b. ? c. !

5. Where are the plants

a. . b. ? c. !

6. Put those matches down

a. . b. ? c. !

B. Add the correct end punctuation and circle which words should be capitalized in each of the following sentences.

1. i am making a hamburger

2. would you like one

3. oh no, I almost burned them

4. please pass the catsup

C. Re-write the following sentences with the correct capitalization and punctuation.

1. i love baked apples

2. does Joe have a dog

Formative Test I Week 3

Grouping Sentences

A. Choose the sentences that are about the same idea and group them together in order. Remember to use right capitalization and punctuation.

1. the teacher gave me a good grade on my homework
2. yesterday I had some homework
3. i rode the bus to school
4. my mother helped me do my school work
5. it took me half hour to do my homework
6. i felt good in class when the teacher checked my homework
7. the weather is getting cold

Corrective I Grouping Sentences Week 3

- A. Circle the letter of the sentences that are about the same idea.
- a. My dog is my best friend.
 - b. He can do many tricks.
 - c. I usually ride my bike to school.
 - d. I take care of him and he takes care of me.
 - e. Today I was dressed up in my nice clothes.
 - f. He is big and fluffy.

Pre Test

Halloween in the United States is a fun holiday for kids. On the lines below write a paragraph about what Halloween means to you and your friends. You may describe the kinds of things people do to have fun, or write about your past experiences and the activities you have planned for this Halloween.

Post Test

Thanksgiving in the United States is a family holiday. On the lines below write a paragraph about what Thanksgiving means to you and your family. You may describe the kinds of things people do to celebrate or write about your past experiences and what your family has planned for this Thanksgiving. If you are from another country and your family does not celebrate Thanksgiving, you may write about what you have learned at school about this holiday.

TABLE OF SPECIFICATIONS PHASE II
Basic Textbook: Different Sources/Teacher Selected Materials

Objectives	Student Behavior					
	know	comp.	apply	anal	synth.	eval.
Part 1.1						
Identification of a Paragraph						
-recognizes Paragraph in text	*	*				
-identifies identified Paragraph -	*	*	*			
identifies one idea Paragraph	*	*	*			
Part 1.2						
Sequencing						
-sequences properly four						
to seven sentence	*	*	*			
-orders sentences in Paragraph						
form	*	*	*			
Part 1.3						
Recognizes a Topic Sentence						
and Content						
-underlines the Topic sentence						
in a sample paragraph	*	*		*		
-chooses the sentence that does						
not belong in a sample						
paragraph	*	*	*			
Part 1.4						
Writing Supporting and Topic						
Sentences						
-writes three supporting						
sentences when given a topic						
sentence	*	*	*			
-writes topic sentence when						
given topic for paragraph	*	*	*			
Part 1.5						
Writing Paragraphs						
-indents two or three sample						
paragraphs	*	*	*			
-writes three paragraphs						
from three sample topic						
sentences	*	*	*			
-writes three paragraphs						
about a topic with three						
parts and at least four	*	*	*			
sentences						

Page 1 CORRECTIVE

Un parrafo es un grupo de frases que indican una idea.
Las ideas deben de presentarse en orden.

Enumera los dibujos y frases en el orden en que tu contaras la accion
secuencial.

Amy se prepara para la carrera.
Ella corre recio.
Ella fue la primera que se preparo.
Ella gano la carrera.

El lo limpio muy bien.
Se veia bien en la mesa.
Jay se encontro un pedazo de madera.
Luego lo pinto.

Formative Test Ia

Un parrafo es un grupo de frases que dicen algo de una idea.
Las ideas deben presentarse en orden.

Estos dibujos estan enumerados en orden.
Enumera en orden las frases que cuentan algo de los dibujos.

[insert pictures]

- ___ Las plantas que se siembran de semilla crecen facilmente.
- ___ Luego pon las semillas cuidadosamente encima.
- ___ Primero pon tierra limpia en una caja de huevos vacia.
- ___ Cubre las semillas con poquita tierra ligeramente.
- ___ Pon la caja en un lugar donde le de sol y riega las semillas.
- ___ Cuando las raices se pongan fuertes, plantalas en tu jardin.
- ___ En algunos dias, chiquitas plantitas verdes empezaran aparecer.

Formative Test Ib

Un parrafo es un grupo de frases que dicen algo de una idea. Las ideas se deben presentarse en orden.

Estos dibujos estan enumerados en orden.

Re-escribe el parrafo que sigue para que las frases esten en el orden correcto.

[insert pictures]

El medico de los ojos dijo que ella necesitaba lentes. Ella fue a ver al medico de los ojos. Karen tenia dificultad con su vista. El le pidio que mirara algunas letras. Karen uso sus lentes nuevos a todas partes. Karen no podia leer las letras pequenas.

Formative Test II

Un parrafo es un grupo de frases que dicen algo de una idea. La frase del tema indica la idea principal del parrafo.

Tom y Ann estan leyendo acerca de gente famosa . Ann tiene un libro sobre Charles Lindberg. El libro de Tom se trata de Amelia Earhart.

Lee cada parrafo que sigue.

Subraya cada frase principal del tema en cada parrafo.

"Primera Dama del Vuelo"

Amelia Earhart se hizo famosa por su amor al vuelo. Ella era una de las primeras mujeres que volo. Ella fue tambien la primer mujer que volo a traves del oceano atlantico.

Amelia Earhart queria volar casi toda su vida. Cuando era mujer joven, tomo muchas lecciones de vuelo. Para el tiempo que cumplio veinte cuatro anos de edad ya podia volar sola. Cuando ella tenia veinte y cinco anos, se compro su primer avion.

Amelia Earhart establecio muchos records. Fue mejor conocida por su viaje transatlantico. Pero, ella fue tambien la primera persona que volo desde Hawaii a America y desde Mexico a Nuevo Jersey.

A menudo Amelia Earhart fue premiada por sus vuelos. Le hacian honor en Francia y los Estados Unidos. Tambien gano otros premios.

En 1937, Amelia Earhart hizo su ultimo viaje. Ella trato de volar alrededor del mundo. En rumbo a casa, su avion bajo. Nunca jamas se supo de ella.

Formative Test III

La frase de tema indica la idea principal del parrafo.
Las otras frases del parrafo deben contar algo sobre la idea principal.

frase del tema

Todos los gatos pertenecen a la misma familia.

Hay gatos pequenos domesticados tambien tigers enormes. Pero todos son buenos cazadores y tienen dientes largos. Mi animal mimado es un perro. **no pertenece**

Lee cada parrafo que sigue.
Sugraya la frase del tema.
Tacha la frase que no pertenece.

Bob fue el mejor jugador en el juego de hoy. El pego tres corridas. Siempre llega tarde a la escuela.
Tambien cogio cuatro pelotas en vuelo.

Hay miles de diferentes clases de estampillas. A Sue se le olvido enviar esta carta. La primera estamilla era de Penny Black de Inglaterra. Fue hecha en 1837. Desde entonces cada pais ha hecho estampillas para su correo.

Tengo que ir a la tienda. Ya no mantequilla. Los trastes no estan limpios.
Tampoco tenemos leche.

Las bicicletas tienen mas de cien anos que existen. Mi bicicleta esta rota. La primera fue construida en 1877. La rueda de enfrente era grande, y la rueda de atras era pequena.

Summative Test I

Todas las frases en un parrafo deben contar sobre una idea principal del parrafo.

Re-escribe cada parrafo que sigue, pero no uses la frase que no pertenece.

Los faros ayudan a las naves a pasar por lugares peligrosos. Al principio la gente usaba lumbres contruidas sobre cerros altos. Naves grades pueden cargar a mucha gente. Ahora los faros usan luces alumbradas para hace el mismo trabajo.

La gente siempre ha usado alguna forma de dinero. El senor Miller es muy rico. Las primeras formas de dinero eran dientes de animales, sal y conchas. Luego aparecieron las monedas. Ahora se usa el dinero en forma de papel.

Los pajaros tienen diferentes modos de cantar. Ellos pueden cantar alto o bajo. Los huevos de pajaros son de diferentes formas. Algunas de sus canciones cuentan de nidos. Otras llamadas canterinas de pajar significan peligro.

corrective

La primera frase del parrafo esta endentada. Para enseñar en donde comienza una nueva idea se escribe en el margen de la izquierda.

_____ La gente primero escribio en piedras, arena y en troncos de arboles. Entonces usaron barro y cueros de animal. Luego hace como mil quinientos anos, la gente del Egipto hizo una forma de papel.

Escribe los siguientes grupos de frases.

Dividelos entre dos parafos.

Ten seguridad de endentar la primera frase de cada parrafo.

Todos los osos tienen cuerpos grandes y colas cortas. Sus piernas son gruesas y fuertes. Ellos andan lentamente. Pero pueden correr 30 millas por hora si tienen alguna razon. El oso mas comun es el oso negro de los Estados Unidos. No siempre es de color negro. Puede ser negro, cafe, gris o casi blanco.

Summative Test II

La primera frase de un parrafo es endentada. La escribimos algunos espacios del margen de la izquierda para enseñar en donde comienza una nueva idea.

Escribe el grupo de frases que siguen.

Dividelas en tres parafos.

Ten seguridad de endentar la primera frase de cada parrafo.

Para hacer el vidrio se calienta una arena especial. El vidrio natural siempre es verde. Pero se le agregan cosas para quitarle el color o para hacer otros colores. Hay muchos modos de hacer algo de vidrio. Pedazos grandes de vidrio se pueden cortar para hacer cosas pequeñas. El vidrio derretido se puede hacer fluir or soplar en formas. Tambien se puede arrollar en hojas largas. Usamos vidrio todos los dias. Algunas cosas que hacemos de vidrio son vasos para tomar, hermosos recipientes de cristal, focos de luz, ventanas, y espejos.

SUMMATIVE TEST III

Escribe este parrafo, pon las frases en el orden corecto.

Suprime la frase que no pertenece.

Subraya la frase del tema.

Hacer diferentes colores con pintura es facil. Luego puedes hacer todos los otros colores simplemente con mezclarlos. Para comenzar debes tener rojo amarillo y azul. Soy orgulloso de mi dibujo.

Escribe el grupo de frases que siguen.

Dividelas en dos parafos.

Ten seguridad de endentar la primera frase de cada parrafo.

El oro es un metal balando y amarillo. Es veinte veces mas pesado que agua. Es facil formarlo y siempre brilla. El oro, se descubrio en California en 1848. Mucha gente se fue alli de- repente. Se querian hacer ricos rapidamente.

Corrective one**EL ORDEN EN UN PARrafo**

Las frases de un parrafo deben ser escritas en el orden en que pasaron los acontecimientos (la accion o las cosas).

Lee las frases que siguen. Luego arreglalas en el orden que ocurren las cosas y entonces escribelas en un parrafo. Las primeras dos frases estan en el orden correcto para ayudarte a comenzar.

1. El cabello de Elena estaba muy largo.
2. Ella decidio ir a que le dieran un corte de pelo.
3. Gloria le dio a Elena un espejo para [insert picture] que viera su nuevo estilo corto de pelo.
4. Cuando Elena llego al salon de belleza, se encontro a Gloria.
5. La madre de Elena hizo una cita en el salon de belleza para el sabado por la manana.
6. Primero Gloria cubrio a Elena con una capa de plastico.
7. Gloria seco y rizo el cabello corto de Elena.

Corrective two**ESCRIBIENDO PARAFOS**

Un parrafo es un grupo de frases que se tratan de una idea.

A. Escribe estas frases en un parrafo. Usa nada mas las frases que se tratan de la misma idea. Endenta la primera linea de el parrafo.

1. Despues de escuela Jorge tiene un trabajo entregando periodicos.
2. Dobla los periodicos y luego los pone en una bolsa.
3. Luego los entriega en su bicicleta.
4. Los amigos de Jorge juegan al futbol despues de escuela.
5. Jorge es un buen nadador.
6. Jorge entriega periodicos a las familias que viven en su cuadra.
7. El deporte favorito de Jorge es beisbal.

[insert picture]

Corrective Three

B. Escribe estas frases en dos parafos. Usa todas las frasesel orden que estan. Asegura que cada parrafo tenga una idea. Endenta la primera linea de cada parrafo.

1. Rita estaba excitada porque iba pasear en avion por primer vez. [insert picture]
2. Cuando estaba esperando en el aeropuerto, ella veia los aviones.
3. Ella vio algunos aviones subir y otros aterrizar.
4. Era tiempo para que Rita se subiera al avion
5. Ella encontro su asiento y se puso el cinturon deseguridad.
6. Ella miro por la ventanilla cuando subia el avion.
7. Todo se veia muy pequeno.

C. En un pliego de papel, escribe un parrafo que se trate de un avion y en el cual a ti te gustaria pasear. Todas las frases deben de ser sobre la misma idea. Endenta la primera linea.

Appendix B

1. Affect Questionnaire English Version
2. Affect Questionnaire Spanish Version

SAMPLE QUESTIONNAIRE

For the following statements **circle** "yes", "don't know" or "no".

Underline the best answer for number 12 and number 13. Think about what it says carefully before you give the final answer. Don't worry about the way you answer because you will not be graded for the answer you give. Try to be as honest as possible about your feelings when you give the final answer.

Part A

- | | | | |
|--|-----|------------|----|
| 1. I like to be called on in my language arts class. | yes | don't know | no |
| 2. I try to do the best work in language arts that I can. | yes | don't know | no |
| 3. My language arts teacher thinks my work is very good. | yes | don't know | no |
| 4. I am very proud of my writing. | yes | don't know | no |
| 5. Writing is easier for me than some of my other subjects. | yes | don't know | no |
| 6. I feel upset in language arts class. | yes | don't know | no |
| 7. I am discouraged with my writing ability. | yes | don't know | no |
| 8. I find it hard to talk in front of my language arts class. | yes | don't know | no |
| 9. Most of the students in my class know more about how to write than I do. | yes | don't know | no |
| 10. My language arts teacher makes me feel that I am doing poorly. | yes | don't know | no |
| 11. I think I am <u>not</u> doing very well in language arts class. | yes | don't know | no |
| 12. What kind of grades do you think you are capable of getting in language arts?
the best grades
average grades
the poorest grades | | | |
| 13. Forget for a minute how others grade your work. How do <u>you</u> think your work is in language arts and particularly in writing?
My work is excellent.
My work is average.
My work is poor. | | | |

Circle "yes", "don't know" or "no" to answer the following statements about your feelings. Remember you will not be graded for the answer that you give but it is very important to answer as honestly as possible.

Part B

- | | | | |
|---|-----|------------|----|
| 1. Learning to write is more difficult to understand than any subject at school. | yes | don't know | no |
| 2. I think I should learn how to write paragraphs. | yes | don't know | no |
| 3. I cannot understand why some students think writing is fun. | yes | don't know | no |
| 4. Learning to write paragraphs is not very useful because it is just putting words together. | yes | don't know | no |
| 5. Writing paragraphs is more like playing a game than school work. | yes | don't know | no |
| 6. Writing paragraphs is boring. | yes | don't know | no |
| 7. I do <u>not</u> think it is important to learn how to write paragraphs. | yes | don't know | no |

Part C

- | | | | |
|---|-----|------------|----|
| 1. Learning to write paragraphs is one of my favorite school subjects. | yes | don't know | no |
| 2. I would like to do more work so that I can learn to write better paragraphs. | yes | don't know | no |
| 3. I would like to show somebody else how to write sentences and paragraphs. | yes | don't know | no |
| 4. I think learning to write sentences and paragraphs is a waste of time. | yes | don't know | no |
| 5. I enjoy learning how to write paragraphs. | yes | don't know | no |
| 6. I want to learn more about writing. | yes | don't know | no |

Spanish Version

Subraya a las siguientes frases "si", "no se" o "no". **Subraya** la mejor respuesta para la pregunta numero 12 y 13. Piensa bien en lo que dice la frase antes the dar la respuesta. No te preocupes de la respuesta que des porque no sera calificada. Trata de dar una respuesta clara a cerca de como te sientes sobre lo que dice la frase.

Parte A

- | | | | |
|--|-------------------|--------------|----|
| 1. A mi me gusta participar en la clase de idioma. | si | no se | no |
| 2. Yo trato de hacer el mejor trabajo posible en la clase de idioma. | si | no se | no |
| 3. Mi maestra de idioma piensa que mi trabajo es bueno. | si | no se | no |
| 4. Yo estoy muy orgulloso/a de my habilidad de escribir. | si | no se | no |
| 5. Escribir es mas facil para mi que algunos de mis otros temas escolares. | si | no se | no |
| 6. Yo me siento trastornado/a en mi clase de idioma. | si | no se | no |
| 7. Yo me siento desanimado/a con mi habilidad de escribir. | si | no se | no |
| 8. Para mi es muy dificil hablar en frente de la clase de idioma. | si | no se | no |
| 9. Casi todos los estudiantes en la clase saben escribir mejor que yo. | si | no se | no |
| 10. Mi maestra de idioma me hace sentir que no estoy haciendo buen trabajo. | si | no se | no |
| 11. Yo pienso que no estoy trabajando bien en mi clase de idioma. | si | no se | no |
| 12. ¿Que grados piensas sacar en la clase de idioma? | | | |
| los mejores grados | grados ordinarios | malos grados | |
| 13. Olvidate por un minuto como califican otros tu trabajo. | | | |
| ¿Como piensas tu calificar tu propio trabajo en la clase de idioma y en tus habilidades de escribir en particular? | | | |
| Mi trabajo es excelente. Mi trabajo es al nivel medio. | | | |
| Mi trabajo es malo. | | | |

Indica "si", "no se" o "no" para responder a las siguientes frases a cerca de como te sientes. Acuerdate que tus respuestas no seran calificadas pero es muy importante que respondas lo mas claro posible.

Parte B

- | | | | |
|---|----|-------|----|
| 1. Aprender a escribir es mas dificil que entender otro tema escolar. | si | no se | no |
| 2. Yo pienso que debo aprender como escribir parafos. | si | no se | no |
| 3. Yo no puedo comprender porque algunos estudiantes piensan que escribir es divertido. | si | no se | no |
| 4. Aprender a escribir parafos no es muy practico porque es nada mas agrupar palabras. | si | no se | no |
| 5. Escribiendo parafos es como jugar un juego, no es como trabajo escolar. | si | no se | no |
| 6. Es aburrido escribir parafos. | si | no se | no |
| 7. Yo no pienso que es importante aprender a escribir parafos. | si | no se | no |

----- Parte C

- | | | | |
|--|----|-------|----|
| 1. Aprender a escribir parafos es uno de mis temas escolares favoritos. | si | no se | no |
| 2. Me gustaria hacer mas trabajo para poder aprender a escribir mejores parafos. | si | no se | no |
| 3. Me gustaria esenarle a otra persona a escribir frases y parafos. | si | no se | no |
| 4. Yo pienso que aprender a escribir frases y parafos es una perdida de tiempo. | si | no se | no |
| 5. Me divierto al aprender a escribir parafos. | si | no se | no |
| 6. Quiero aprender mas para poder escribir mejor. | si | no se | no |

Appendix C

1. CSE Subscales Scoring Categories
2. Trait Scales Scoring Categories

CENTER FOR THE STUDY OF EVALUATION

Expository Scale IV

Dr. Edys Quallmalz, principal author

CSE SUBSCALES 1,2,4,6

(Subscales used for the present study)

****Note:** some views on exposition;

Exposition is the kind of discourse that explains or clarifies a subject.

Exposition seeks to explain or inform through such methods as giving reasons or examples, comparing and contrasting, defining, enumerating, or, through a combination of methods.

Exposition explains why or how.

Exposition promotes reader understanding on a subject.

Subscale One : General Impression

You are to read each paragraph quickly first, in order to form an overall impression of its quality. To assign the paragraph a score, consider the following question: To what extent does the paragraph achieve an expository purpose for the intended audience?

Master

6 = An excellent example of exposition

5 = A good example of exposition

4 = An adequate example of exposition

Non-master

3 = A marginal example

2 = A poor example

1 = A very poor example, or barely readable
the topic

Subscale Two: General Competence

Based on your first or second reading of the paragraph, decide how competently the writer formed the paragraph. Does the paragraph demonstrate mastery or command of just the basic paragraph elements listed below? If the student received no further writing instruction do you think he or she would produce other writing which communicates clearly and exhibits command of these elements: main idea; paragraph organization, support, mechanics (usage, sentence construction, spelling, punctuation, capitalization)?

Master6 = Very Competent

The paper executes all the elements competently. There are no serious errors. The paper has a clear main idea, logical organization, relevant, detailed support, and a command of basic mechanics. There are no major flaws.

5 = Definitely Competent

The paper is competent in all of the basic elements, but there may be a few minor flaws.

4 = Adequately Competent

The paper is adequately competent in all elements. There may be some serious flaws.

Non-Master3 = Almost Competent

The paper lacks competence in one or two elements, and there are a few major flaws.

2 = Not Very Competent

The paper has two or more of the elements under control. There are many serious flaws.

1 = Not At All Competent

The paper has none or only one of the elements executed competently.

Subscale Four: Paragraph Coherence

This subscale focuses upon the relationship of ideas within a paragraph, their logical interrelationship and subordination to the paragraph topic. Ideally, the paragraph presents subtopics which are developed by cohesive groups of supporting statements. Each paragraph represents a complete unit of thought. Statements within the paragraph relate logically to each other and to the paragraph topic.

MasterSix

All major units of thought appear in the paragraph. The paragraph has a clearly stated or implied topic. All sentences within the paragraph are related to each other, to the paragraph topic and are subordinate to it. There are no one-sentence paragraphs, unless they are especially effective.

Five

The topic is developed in the paragraph. The paragraph contains logically related subordinate support. There may be a minor digression.

Four

The topic is developed in discrete sentences but are somehow related. There may be some minor digressions.

Non-MasterThree

The paragraph statements are not logically related and do not function as subordinate support to the paragraph topic. Some relationships between sentences must be inferred.

Two

There are few statements that are logically related or supported. There are many digressions. Many relationships among sentences must be inferred.

One

There are no statements that logically cohere.

Subscale Six: Mechanics

This category covers usage, sentence construction, spelling, punctuation, and capitalization. Following the descriptions of each score category is a reference list of "errors" of "flaws" considered serious.

MasterSix

There are few or no errors. There are no serious errors.

Five

There may be a few minor errors in the usage, sentence construction, spelling, punctuation, or capitalization, but no more than one serious error.

Four

There are some errors in the Mechanics categories. A few may be serious.

Non-MasterThree

There are numerous errors in the categories. There are some serious errors in several categories. Students' sentence construction is judged below mastery.

Two

There are many serious errors, causing the reader some confusion.

One

Errors are so numerous and serious that they interfere with communication.

Error Categories and Examples of Serious Errors

1. Sentence Construction: subject-verb agreement, run-on and fragment sentences.
2. Usage: homonyms (its, it's; their, there), incorrect use of common words, incorrect pronoun referents
3. Spelling: common words misspelled. Note: any misspelled word only counts as one error, even if the misspelling is repeated.
4. Punctuation/Capitalization: Contractions, commas, sentence punctuation.
5. Paragraph Conventions: titled or numbered paragraphs

Description of Primary Trait

This primary trait focuses upon the relationship of ideas within a paragraph and their logical interrelationship and subordination to a topic or main idea sentence. Ideally, the topic or main idea sentence states explicitly the idea that unifies all subsequent sentences (a component of unified thought). Sentences within the paragraph relate logically to each other and to the paragraph topic. Each paragraph represents a complete unit of thought.

Mastery

Five

The paragraph receiving a score of five has an explicit topic sentence to which everything else is subordinated. Sentences must elaborate by using description, examples, or detail.

Four

Four has an explicit topic sentence. There may be some minor digressions in thought but generally this paragraph has some unity and sentences have a logical sequence.

Non Mastery

Three

Paragraph receiving a score of three has a topic sentence which is implicit but not clearly stated. Sentences logically relate to each other and support the main idea.

Two

Paragraphs receiving a score of two do not have a topic sentence. There may be two or more sentences that relate to each other in content but do not work to support each other (one sentence is not an elaboration of the other nor does it support the other in detail).

One

One or more sentences that do not relate to the assigned topic.

Secondary Trait
Description

Inventive expression and elaboration on a topic.

Rationale

The rationale of the secondary trait is to test whether a student can draw upon his or her own experiences or use his or her imagination to respond to a writing prompt. The writer can project, invent or remember experiences in or outside of the school or family setting. The writer must select information i.e., details, description etc. , from these experiences or settings that will provide consistency in his or her response to the topic. All topics ask the writer to consider a personal experience whether imaginative or real. The selection and ordering of detail for purposes of elaboration should make the best papers highly structured, thus these papers are coherent. Structure, in this sense, does not define content, but rather content defines structure. Weak writers will not provide details, examples, descriptive language or words i.e., adjectives or adverbs, to elaborate on a topic. Strong well elaborated responses will develop a topic by using such devices as dialogue, end punctuation marks, titles and may go beyond merely listing details.

Mastery

Five

Level five papers provide numerous details, create abstraction and strong devices, (i.e., dialogue, end punctuation marks, titles) to elaborate on a topic. The selection and ordering of details for purposes of elaboration make the papers at this level highly structured. The inventive expression is produced by inventing abstraction through hypothesizing and imagery.

Four

Level four show the writer's use of experience or imagination to express concrete thoughts. The distinction between the response which receives a four and one which receives a five is the level of abstraction expressed through the description. The four paper responds to a topic by projecting feeling, listing descriptive detail, remembering past experience and providing enough detail to elaborate on the topic.

But elaboration is based on concrete thought; i.e. some reality based statement rather than fantasy or imagery. Content should define a sense of structure, or continuity by elaborating. In addition, content is planned and shows intended purpose.

Non-Mastery

Three

Level three writers use their experience or imagination to respond to the topic at a concrete level. Excessive details are used inappropriately and hinder elaboration by not providing continuity. Deviations in thought create gaps and unevenness in the elaboration of ideas. Since elaboration is flawed, and details may be merely named instead of described, structure is inadequate.

Two

Writer does not draw on his or her past experience in or outside of school or family to develop a topic. Response has limited elaboration which may be disjointed. Information is inappropriately presented. There is no clear purpose or plan in writing. There is an attempt to write on a topic but descriptions are all at the concrete level.

One

These papers clearly show incomplete thoughts. They may list within a structure but the list is merely listing. The response does not provide elaboration for a consistent unified thought. There is no attempt to express an idea or more than one idea.